06-15-06

AR

OIPE IAP92 JUN 14 2006 PATENT & TRADEMARK OFFICE

Express Mail Mailing Label No. EV832482014US

TRANSMITTAL FORM

Application Serial Number	09/912,947
Filing Date	July 25, 2001
First Named Inventor	Dahlbäck
Group Art Unit	1634
Examiner Name	Bausch, Sarae L.
Attorney Docket No.	INL-036DV
Patent No.	Not applicable
Issue Date	Not applicable

ENCLOSURES *(check all that apply)*

- ☒ Fee Transmittal Form
  - ☒ Check Attached
  - ☐ Copy of Fee Transmittal Form
- ☐ Amendment/Response
  - ☐ Preliminary
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
  - ☐ Letter to Official Draftsperson including Drawings [Total Sheets _____]
- ☐ Petition for Extension of Time
- ☐ Information Disclosure Statement
  - ☐ Form PTO-1449
  - ☐ Copies of IDS Citations
- ☐ Certified Copy of Priority Document(s)
- ☐ Sequence Listing submission
  - ☐ Paper Copy/CD
  - ☐ Computer Readable Copy
  - ☐ Statement verifying identity of above
- ☐ Copy of Notice to File Missing Parts of Application
- ☐ Formal Drawing(s)
- ☐ Request For Continued Examination (RCE) Transmittal
- ☐ Power of Attorney (Revocation of Prior Powers)
- ☐ Terminal Disclaimer
- ☐ Executed Declaration and Power of Attorney for Utility or Design Patent Application
- ☐ Small Entity Statement
- ☐ CD(s) for large table or computer program
- ☐ Amendment After Allowance
- ☐ Request for Certificate of Correction
  - ☐ Certificate of Correction (in duplicate)
- ☐ Notice of Appeal to Board of Patent Appeals and Interferences
- ☒ Appeal Brief
  - ☒ Claims Appendix
  - ☒ Evidence Appendix A-H
  - ☒ Related Proceedings Appendix
- ☐ Status Inquiry
- ☒ Return Receipt Postcard
- ☐ Certificate of First Class Mailing under 37 C.F.R. 1.8
- ☐ Certificate of Facsimile Transmission under 37 C.F.R. 1.8
- ☐ Additional Enclosure(s) *(please identify below)*

CORRESPONDENCE ADDRESS

Direct all correspondence to:
Patent Administrator
Kirkpatrick & Lockhart Nicholson Graham LLP
State Street Financial Center
One Lincoln Street
Boston, MA 02111-2950
Tel. No.: (617) 261-3100
Fax No.: (617) 261-3175

SIGNATURE BLOCK

Respectfully submitted,

Date: June 14, 2006
Reg. No. 51,551
Tel. No.: (617) 261-3198
Fax No.: (617) 261-3175

Fangli Chen, Ph.D.
Agent for Applicant
Kirkpatrick & Lockhart Nicholson Graham LLP
One Lincoln Street
Boston, MA 02111-2950

Express Mail Mailing Label No. EV832482014US

FEE TRANSMITTAL FY 2006

	Complete if Known
Application Serial Number	09/912,947
Filing Date	July 25, 2001
First Named Inventor	Dahlbäck
Group Art Unit	1634
Examiner Name	Bausch, Sarae L.
Attorney Docket No.	INL-036DV

METHOD OF PAYMENT

1. ☒ Payment Enclosed:
   ☒ Check ☐ Money Order ☐ Other

2. ☒ The Commissioner is hereby authorized to credit or charge any fee indicated below for this submission to Deposit Account No. 50-1721.
   ☐ Required Fees (copy of this sheet enclosed).
   ☒ Additional fee required under 37 CFR 1.16 and 1.17.
   ☒ Overpayment Credit.

3. ☐ Applicant claims small entity status.

FEE CALCULATION

1. FILING/SEARCH/EXAM/SIZE FEES

Large Entity Fee ($)	Fee Description	Fee Paid
300	Utility filing fee	
500	Utility search fee	
200	Utility exam fee	
250	Utility size fee (each add'l 50 pgs. over 100)	
200	Design filing fee	
100	Design search fee	
130	Design exam fee	
250	Design size fee (each add'l 50 pgs. over 100)	

	Number Filed	Number Extra	Rate	Amount
Total Claims		- 20 =	x $ 50.00 =	
Independent Claims		- 3 =	x $200.00 =	
☐ Multiple Dependent Claim(s), if any			$360.00 =	
			TOTAL:	
			SMALL ENTITY DISCOUNT:	
			SUBTOTAL (1)	($) 0.00

2. AMENDMENT CLAIM FEES

	Claims Remaining After Amend.	Highest No. Previously Paid For	Present Extra	Rate	Fee Paid
Total	-	=		x $ 50.00 =	
Indep.	-	=		x $ 200.00 =	
☐ First Presentation of Multiple Dep. Claim				+ $ 360.00 =	
				TOTAL:	($)
				SMALL ENTITY DISCOUNT:	($)
				SUBTOTAL (2)	($)0.00

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee ($)	Small Entity Fee ($)	Fee Description	Fee Paid
130	65	Surcharge - late filing fee or oath	
50	25	Surcharge - late provisional filing fee or cover sheet	
130	130	Non-English specification	
2,520	2,520	Request for ex parte reexamination	
120	60	Extension for reply within first month	
450	225	Extension for reply within second month	
1020	510	Extension for reply within third month	
1590	795	Extension for reply within fourth month	
2160	1080	Extension for reply within fifth month	
500	250	Notice of Appeal	
500	250	Filing a brief in support of an appeal	500.00
1000	500	Request for oral hearing	
400	400	Petitions to the Commissioner (Gp. I)	
200	200	Petitions to the Commissioner (Gp. II)	
130	130	Petitions to the Commissioner (Gp. III)	
180	180	Submission of Information Disclosure Statement	
790	395	Filing a submission after final rejection (37 CFR 1.129(a))	
790	395	For each additional invention to be examined (37 CFR 1.129(b))	
100	100	Certificate of Correction for applicant's error	
130	65	Submission of Terminal Disclaimer	
Other fee (Specify)			
Other fee (Specify)			

SUBTOTAL (3) ($) 500.00

SUBTOTAL (1)	0.00
SUBTOTAL (2)	0.00
SUBTOTAL (3)	500.00
TOTAL	($) 500.00

CORRESPONDENCE ADDRESS

Direct all correspondence to:
Patent Administrator
Kirkpatrick & Lockhart
Nicholson Graham LLP
State Street Financial Center
One Lincoln Street
Boston, MA 02111-2950
Tel. No.: (617) 261-3100
Fax No.: (617) 261-3175

SIGNATURE BLOCK

Respectfully submitted,

Date: June 14, 2006
Reg. No. 51,551
Tel. No.: (617) 261-3198
Fax No.: (617) 261-3175

Fangli Chen, Ph.D.
Agent for Applicant
Kirkpatrick & Lockhart Nicholson Graham LLP
One Lincoln Street
Boston, MA 02111-2950

Express Mail Mailing Label No.: EV832482014US

PATENT

Attorney Docket No. INL-036DV

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPELLANT(S):	Dahlbäck		
SERIAL NO.:	09/912,947	GROUP NO.:	1634
FILING DATE:	July 25, 2001	EXAMINER:	Bausch, Sarae L.
TITLE:	*Assays for Determining Anticoagulant Cofactor Activity*		

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

BRIEF IN SUPPORT OF AN APPEAL

Sir:

This is an appeal from the final rejection of claims 46, 53, 54, 55, 64 and 65 by the Examiner set forth in the Final Office Action, mailed from the U.S. Patent and Trademark Office (USPTO) on December 27, 2005.

This Appeal Brief is submitted pursuant to the Notice of Appeal filed with the USPTO on April 24, 2006. Appellant includes a check for $500.00 to cover the fee for filing a brief in support of an appeal according to 37 C.F.R. 41.20(b)(2). Appellant believes no other fees are necessary for consideration of this paper. However, if a further fee is required, please consider this a conditional petition therefor and authorization to charge Deposit Account No.50-1721.

I. REAL PARTY IN INTEREST

Named inventor, Björn Dahlbäck, transferred the entire right, title and interest of the present application to T.A.C. Thrombosis and Coagulation AB by virtue of an assignment on March 22, 1997. A copy of the assignment was filed in the parent of the present application, U.S. Serial No. 08/500,917, and recorded at the U.S. Patent and Trademark Office on April 7, 1997, at Reel 8442, Frame 0596. A copy of the assignment was also filed in the present application on June 8, 2006.

T.A.C. Thrombosis and Coagulation AB entered an exclusive licensing agreement with Instrumentation Laboratory S.p.A. for and to the technology disclosed in the present application. An advisement of this license was recorded at the U.S. Patent and Trademark Office on April 28, 1997, at Reel 8470, Frame 0595.

Thus, the real parties in interest with regard to the present application are T.A.C. Thrombosis and Coagulation AB and Instrumentation Laboratory S.p.A.

II. RELATED APPEALS AND INTERFERENCES

Appellant wishes to make the Board aware that the parent of the present application, U.S. Serial No. 08/500,917, is undergoing the following interference proceedings.

Patent Interference No. 105,235	Griffin *et al.* (Patent 5,705,395) v. Dahlbäck (USSN 08/500,917)
Patent Interference No. 105,268	Griffin *et al.* (Patent 5,834,223) v. Dahlbäck (USSN 08/500,917)
Patent Interference No. 105,269	Griffin *et al.* (Patent 6,083,757) v. Dahlbäck (USSN 08/500,917)

Appellant is not aware that any decisions have been rendered by the Board in any one of the proceedings identified above.

III. STATUS OF THE CLAIMS

Claims 1-43 were originally filed in this application. In the Preliminary Amendment filed on July 25, 2001, and entered by the Examiner, Appellant cancelled claims 1-39 and 43, amended claims 40, 41 and 42, and introduced new claims 44-52. In the Amendment and Response filed on April 9, 2004, and entered by the Examiner, claims 40 and 46-52 were amended, and new claims 53-61 were introduced. In the Amendment and Response filed on December 8, 2004, and entered by the Examiner, claims 40-42, 44, 45, 47, 48 and 50-52 were cancelled, claims 46, 49, 53, 54, 57, 59 and 60 were amended, and new claims 62 and 63 were introduced. In the Amendment and Response filed on September 28, 2005, and entered by the Examiner, claims 49 and 56-63 were cancelled, claims 46 and 53-55 were amended, and new claims 64 and 65 were introduced. Appellant therefore appeals claims 46, 53-55, 64 and 65. The claims on appeal are provided in the attached Claims Appendix.

IV. STATUS OF AMENDMENTS

No amendment was filed subsequent to final rejection of claims 46, 53, 54, 55, 64 and 65, in the Final Office Action mailed from the USPTO on December 27, 2005.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

The core of the present invention is the discovery that Factor V has a novel anticoagulant activity (also referred to as APC-cofactor 2 activity in the specification). See, *e.g.*, the original specification, page 3, line 21, to page 4, line 15 (*i.e.*, paragraph 0018 of the substitute specification filed on March 6, 2002, and entered by the Examiner). Throughout the specification, Appellant provides extensive clinical and biochemical evidence supporting the conclusion that Factor V has a novel anticoagulant activity and that deficiency of such activity causes thrombosis associated with Activated Protein C (APC)-resistance. See, *e.g.*, the original specification, page 6, line 4, to page 7, line 12 (*i.e.*, the substitute specification, paragraph 0027).

Claims directed to functional assays of the anticoagulant activity of Factor V and the use of the functional assays to diagnose blood coagulation/anticoagulation disorders (*e.g.*, thrombosis) have been allowed in the parent application, U.S. Serial No. 08/500,917.

The present application contains two independent claims, claims 46 and 54. Independent claim 46 is directed to a method for detecting an individual at risk of developing thrombosis. Claim 46 requires obtaining a sample from an individual, conducting a nucleic acid assay such as a hybridization or sequencing assay, and determining abnormal presence or absence of at least one nucleic acid fragment or sequence in the individual's Factor V gene compared to a normal control.

Support for claim 46 can be found in the specification as originally filed at least, for example, at page 20, lines 12-19 (*i.e.*, the substitute specification, paragraph 0081).

Independent claim 54 is directed to a method for determining a presence of a Factor V gene mutation associated with APC-resistance in an individual at risk for APC-resistance. Claim 54 requires obtaining a sample from an individual at risk for APC-resistance, conducting a nucleic acid sequencing assay on the sample using reagents specific for the Factor V gene to determine the Factor V gene sequence, and determining the presence of the Factor V gene mutation associated with APC-resistance in the individual by comparing the determined Factor V gene sequence to a normal Factor V gene sequence. The subject matter of claim 54 differs from that of claim 46 at least because claim 54 requires obtaining a sample from an individual known to have APC-resistance or to be at risk for APC-resistance.

Support for claim 54 can be found in the specification as originally filed at least, for example, at page 20, lines 12-23 (*i.e.*, the substitute specification, paragraph 0081).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

One ground of rejection raised by the Examiner in the Final Office Action is to be reviewed on appeal. That is the rejection of claims 46, 53-55, 64 and 65 under 35 U.S.C. § 112, first paragraph, as failing to comply with the enablement requirement.

VII. APPELLANT'S ARGUMENT

Claims 46, 53-55, 64 and 65 stand rejected under 35 U.S.C. § 112, first paragraph, for allegedly failing to enable one skilled in the art to make and/or use the invention without undue experimentation. Specifically, the Examiner alleges that the nature of the claimed invention requires the knowledge of a specific mutation in the Factor V gene associated with the increased

risk of thrombosis or APC-resistance. The specification does not teach such a specific mutation. Therefore, according to the Examiner, the quantity of experimentation necessary to practice the claimed invention is very high and would involve the screening and analysis of the Factor V gene from hundreds of patients to identify any putative mutations within the gene to establish a relationship between the mutations and the recited phenotypes. See, the Final Office Action, pages 3, 4 and 5. The Examiner asserts that although the specification teaches that a neutral polymorphism in the Factor V gene has linkage with inherited APC-resistance, the specification does not disclose the nature of the polymorphism that result in the expression of APC-resistance (see, Final Office Action, page 4). In raising this rejection, the Examiner relied on the examples discussed in Pennisi, Science, 281 (5384):1787-1789, and the heterozygous single base polymorphisms (*i.e.*, biallelic markers) in the FLAP gene described in WO 99/52942 to show that it is highly unpredictable as to whether a particular polymorphism marker will be associated with a particular disease (see, Office Action, pages 4 and 5). A copy of Pennisi is enclosed as Evidence Appendix A. A copy of WO 99/52942 is enclosed as Evidence Appendix B.

The test for enablement is whether one reasonably skilled in the art could make or use the invention as broadly as it is claimed based on the disclosures in the specification coupled with information known in the art without undue experimentation. See *In re* Wands, 858 F.2d 731 (CAFC 1988). In Wands, the court faced the question whether the specification of the Wands patent enabled one skilled in the art to make high affinity IgM monoclonal antibodies for hepatitis B-surface antigen. The Wands court recognized that the nature of monoclonal antibody technology involved screening hybridomas to determine which hybridomas secrete antibodies with desired characteristics. Id. at 740. The court stated: "Enablement is not precluded by the necessity for some experimentation such as routine screening. However, experimentation needed to practice the invention must not be undue experimentation. 'The key word is 'undue,' not 'experimentation.'" Id. at 736-737. "The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed." Id. at 737. In deciding whether undue experimentation is involved for practicing the invention as claimed, the court considered the following eight factors: "(1) the quantity of experimentation necessary, (2) the amount of direction or guidance

presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the relative skill of those in the art, (7) the predictability of the art, and (8) the breadth of the claims." Id. at 737. The court concluded that undue experimentation would not be required to practice the invention because (1) Wands' disclosure provided considerable direction and guidance on how to practice the invention, (2) there was a high level of skill in the art at the time when the application was filed, and (3) all of the methods needed to practice the invention were well known. Id. at 737, 740.

Recently, the United States Court of Appeals for the Federal Circuit further clarified the purpose of the enablement requirement in its decision in Invitrogen Corporation v. Clontech Laboratories, Inc., 429 F.3d 1052 (CAFC, 2005). The court stated that: "Section 112 requires that the patent specification enable 'those skilled in the art to make and use the full scope of the claimed invention without 'undue experimentation' in order to extract meaningful disclosure of the invention and, by this disclosure, advance the technical arts." Id. The court further stated that: "The scope of patent claims must be less than or equal to the scope of the enablement. The scope of enablement, in turn, is that which is disclosed in the specification plus the scope of what would be known to one of ordinary skill in the art without undue experimentation."

Claims 46, 53, 64 and 65

Independent claim 46 is summarized in Section V above. The full text of claim 46 is provided in full in the Claims Appendix that follows. Claim 46 is directed to a method for detecting an individual at risk of developing thrombosis by detecting abnormal presence or absence of at least one nucleic acid fragment or sequence in the individual's Factor V gene compared to a normal control.

Appellant submits that the present application is distinguishable from Pennisi and WO 99/52942 relied on by the Examiner for the following reasons. Numerous methods may be used to associate disease genes with disease phenotypes. For example, Pennisi and WO 99/52942 use genomic polymorphisms to associate disease genes with disease phenotypes. Specifically, Pennisi discusses the general unpredictability in using genomic polymorphisms (*e.g.*, SNPs) to associate disease genes with disease phenotypes. WO 99/52942 describes heterozygous single base polymorphisms (*i.e.*, biallelic markers) in the FLAP gene and the potential association of

these markers to certain diseases such as asthma. Neither Pennisi nor WO 99/52942 discusses yet another method for associating disease genes with disease phenotypes, *i.e.*, using biochemical data. By contrast, the present application conclusively associated Factor V with thrombosis or APC-resistance using extensive biochemical and clinical data without relying on the polymorphism in the Factor V gene.

As summarized in Section V above, the core of the present invention is the discovery that Factor V has a novel anticoagulant activity (also referred to as APC-cofactor 2 activity in the specification). See, *e.g.*, the original specification, page 4, lines 2-4 (*i.e.*, the substitute specification, paragraph 0019). Throughout the specification, Appellant provides extensive clinical and biochemical evidence leading to the conclusion that Factor V has a novel anticoagulant activity and that deficiency of such activity causes thrombosis associated with APC-resistance. For example, as set forth on page 3, line 21, to page 4, line 15, the specification teaches the following:

> "According to the present invention it has been found that APC-resistance is due to deficiency of a previously unrecognized anticoagulant cofactor activity enhancing the proteolytic effect of APC directed against Factor Va and Factor VIIIa. . . . More specifically, this anticoagulant activity has been found to be expressed by Factor V, a finding that is quite surprising, since Factor V is the precursor to the procoagulant Factor Va, the latter being degraded by APC in the above mentioned Protein C anticoagulant system. . . . The discovery of the novel anticoagulant cofactor activity according to the present invention is based on the discovery of one patient with thrombosis and an abnormal APC-resistance when his plasma was assayed with the methods disclosed in WO 93/10261 (incorporated herein by reference) and by Dahlbäck et al. (Thromb. Haemost. 65, Abstract 39 (1991) 658)." [Emphasis added. See, the substitute specification, paragraphs 0018, 0019 and 0020.]

On page 6, line 4, to page 7, line 12, the specification discloses 9 different categories of biochemical evidence supporting the conclusion that Factor V has a novel anticoagulant activity (*i.e.*, APC-cofactor 2 activity). In particular, the specification discloses that the anticoagulant activity co-migrates with Factor V protein on SDS-PAGE (category 1); that antibodies specific against human Factor V protein react with the three protein bands associated with the anticoagulant activity in Western blotting (category 2); that monoclonal antibodies obtained

using preparations containing the anticoagulant activity react with Factor V protein (category 4); and that Factor V and the anticoagulant activity coelute on every chromatography (category 5).

Therefore, unlike the examples described in Pennisi and WO 99/52942 which primarily relied on genomic polymorphisms to associate disease genes with disease phenotypes, the present application has conclusively associated Factor V with thrombosis/APC-resistance based on extensive biochemical and clinical evidence without relying on the polymorphisms in the Factor V gene. The linkage of a neutral polymorphism in the Factor V gene and the expression of APC-resistance disclosed on page 20, lines 8-23, of the original specification, only serves as one additional piece of evidence that corroborates with the extensive biochemical and clinical data disclosed in the specification. The present disclosure and claimed invention, however, does not rely on the nature of this polymorphism to establish the association between Factor V and thrombosis/APC-resistance. Therefore, Appellant submits that the general unpredictability in using genomic polymorphisms to associate disease genes with disease phenotypes does not apply to the present disclosure and claimed invention.

Appellant further submits that the knowledge of a specific mutation in the Factor V gene is not required to practice the invention as recited in claim 46. Claim 46 is directed to a method for detecting an individual at risk of developing thrombosis. In another words, claim 46 only requires detecting an individual who is more likely to develop thrombosis than a normal individual. Appellant submits that claim 46 does not require detecting a specific mutation in an individual's Factor V gene in order to determine that the individual is more likely to develop thrombosis than a normal individual. Detection of a mutation manifested by at least one abnormal nucleic acid fragment or sequence in an individual's Factor V gene would be sufficient to indicate that the individual is more likely to develop thrombosis than a normal individual because the Factor V gene containing an abnormal nucleic acid fragment or sequence is more likely to have an abnormal anticoagulant activity compared to a normal Factor V gene. Similar correlation between defects in genes without knowledge of the specific mutation and disease is known in the medical arts. For example, it has been well accepted that an individual is more likely to develop cancer than a normal individual if the individual's p53 gene (tumor suppressor gene) contains a mutation (*i.e.*, abnormal nucleic acid fragment or sequence) compared to a normal control, even if the specific nature of the mutation is unknown.

Therefore, Appellant submits that the nature of the invention as claimed in claim 46 does not require the knowledge of a specific mutation. It only requires determination of abnormal presence or absence of at least one nucleic acid fragment or sequence in an individual's Factor V gene compared to a normal control as recited in claim 46.

Applicant further submits that the present specification fully enables one of skill in the art to determine abnormal presence or absence of at least one nucleic acid fragment or sequence in an individual's Factor V gene compared to a normal control as recited in claim 46. First, like in Wands, the specification of the present application provides reasonable guidance or directions on how to practice the claimed invention. For example, as set forth on page 20, lines 8-23, the specification teaches the following:

> "This is conclusive evidence that nucleic acid hybridisation assays, as well as nucleic acid sequencing can be used in conventional ways in order to detect individuals at risk for thrombotic events due to a low level of APC-cofactor 2 activity. Thus, these types of assays may be used for checking, in an individual, the abnormal presence or absence of one or more nucleic acid fragment(s) and/or sequence(s) unique for the presence or absence of expression of a Factor V molecule either carrying APC-cofactor 2 activity or being deficient in this activity. The protocols and conditions are the same as normally applied for other genes, except for now using reagents specific for the Factor V gene and, optionally, mutation(s) associated with APC-resistance or specific for a normal Factor V gene. Any cell sample from the individual may be appropriate." [See, the substitute specification, paragraph 0081.]

One of skill in the art upon review of the above paragraph would readily have understood how to carry out nucleic acid assays, such as hybridization or sequencing assays, using reagents specific to the Factor V gene and to compare the hybridization or sequencing result to a normal control in order to determine abnormal presence or absence of at least one nucleic acid fragment or sequence in the Factor V gene.

Secondly, like in Wands, there was a high level of skill in the art of Factor V gene and protein in 1993 when the priority applications (Swedish Application Nos. 9302457-8 and 9300300-2) were filed. Specifically, both the Factor V protein sequence and the Factor V cDNA sequence were known in the art before 1993. For example, Jenny *et al.* and Kane *et al.* reported complete cDNA and amino acid sequences of human Factor V in 1987. Jenny J.R. *et al.*, 1987, "Complete cDNA and derived amino acid sequence of human factor V," PNAS USA, 84:4846-

4850, a copy of which is enclosed as Evidence Appendix C. Kane W.H. *et al.*, 1987, "Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin," PNAS USA, 83:6800-6804, a copy of which is enclosed as Evidence Appendix D. In addition, Cripe *et al.* reported the genomic structure of the human Factor V gene in 1992. Cripe L.D. *et al.*, 1992, "Structure of the Gene for Human Coagulation Factor V," Biochemistry, 31:3777-3785, a copy of which is enclosed as Evidence Appendix E. Evidence Appendixes C, D and E were submitted together with the Amendment and Response filed on September 28, 2005, and were entered in the record by the Examiner.

Appellant further submits that, like in Wands, all of the methods needed to detect abnormal presence or absence of a nucleic acid fragment or sequence in the Factor V gene were well known in the art when the application was filed. In the art of molecular biology when this application was filed, it was routine for one of ordinary skill in the art to isolate nucleic acids from a cell sample, to conduct a nucleic acid assay such as hybridization or sequencing, to determine nucleic acid or protein fragment patterns or sequences, and to compare the determined patterns or sequences to normal controls in order to detect a presence or absence of abnormal fragment or sequence. For example, general methods and tools were described in Alberts B. *et al.* (eds.), 1986, "Molecular Biology of the Cell," Garland Publishing, Inc., New York & London, pp 185-196, a copy of which is enclosed as Evidence Appendix F. Methods and tools (*e.g.*, primers and other reagents) specific for the isolation, amplification and sequencing of the Factor V cDNA directly from human tissues, *e.g.*, lymphocytes, were disclosed in Shen *et al.*, April 1, 1993, "The Serine Protease Cofactor V Is Synthesized by Lymphocytes," J. Immunology, 150:2992-3001, a copy of which is enclosed as Evidence Appendix G. Evidence Appendixes F and G were submitted together with the Amendment and Response filed on September 28, 2005, and were entered in the record by the Examiner.

Furthermore, it was well within routine skills of an ordinary artisan when the application was filed to determine what sequence substitution in the Factor V gene would be silent or cause only conservative substitution and what sequence substitution would alter Factor V's function resulting in increased risk of developing thrombosis. For example, similar or conservative amino acids in general were described in Dayoff *et al.*, 1978, Atlas of Protein Sequence and Structure, M. O. Dayoff, ed., Nat'l Biomed. Research Fnd., Washington D.C., vol. 5, Suppl. 3, pp. 345-362.

With respect to Factor V, Shen *et al.* discussed in detail that there are certain mutations in the Factor V gene that are silent or only cause conservative substitutions and may not significantly or only subtly affect the Factor V protein function. See, page 2996, left column of Evidence Appendix G.

Thus, in view of the teachings or the present application and armed with the knowledge available in the art, one of ordinary skill would readily have been able to detect an individual who is more likely to develop thrombosis than a normal individual by determining abnormal presence or absence of at least one nucleic acid fragment or sequence in the individual's Factor V gene compared to a normal control.

For at least the above reasons, Appellant respectfully submits that the present application fully complies with the enablement requirement with respect to independent claim 46 and its dependent claims 53, 64 and 65 and request reconsideration and withdrawal of the rejection of claims 46, 53, 64 and 65 under 35 U.S.C. § 112, first paragraph.

Claims 54 and 55

Independent claim 54 is summarized in Section V above. The full text of claim 54 is provided in full in the Claim Appendix that follows. Claim 54 relates to a method for determining a presence of Factor V gene mutation associated with APC-resistance in an individual at risk for APC-resistance. Claim 54 requires obtaining a sample from an individual known to have APC-resistance or be at risk for APC-resistance, conducting a nucleic acid sequencing assay on the sample using reagents specific for the Factor V gene to determine the Factor V gene sequence and determining the presence of the Factor V gene mutation associated with APC-resistance in the individual by comparing the determined sequence to a normal Factor V gene sequence.

Appellant submits that the knowledge of specific mutation in the Factor V gene is not required to practice the invention as claimed in claim 54. As argued above with respect to claims 46, 53, 64 and 65, the present application has conclusively established the association between Factor V and APC-resistance based on extensive clinical and biochemical evidence. See, *e.g.*, the original specification, page 3, line 21, to page 4, line 12; and page 6, line 4, to page 7, line 12 (*i.e.*, the substitute specification, paragraphs 0018, 0019 and 0020). Since claim 54 requires

obtaining a sample from an individual known to have or to be at risk for APC-resistance, the mutation determined in the individual's Factor V gene by conducting a nucleic acid sequencing assay on the sample as recited in claim 54 would naturally be associated with APC-resistance. Therefore, Appellant submits that the nature of the claimed invention does not require the knowledge of a specific mutation in the Factor V gene. It would also not be necessary to screen and analyze the Factor V gene from hundreds of patients to identify any putative mutations within the gene to establish a relationship between the mutations and the recited phenotype, *i.e.*, APC-resistance, in order to practice the invention as claimed in claim 54.

Appellant submits that, based on the teachings of the present application and the knowledge available in the art, one of skill in the art would readily have been able to practice the invention as claimed in claim 54 with only routine experimentation. Like in Wands, the specification of the present application provides reasonable guidance or directions on how to conduct a nucleic acid sequencing assay in an individual's Factor V gene using reagents specific for the Factor V gene as recited in claim 54. See, *e.g.*, the original specification, page 20, lines 8-23 (*i.e.*, the substitute specification, paragraph 0081).

As discussed above with respect to claims 46, 53, 64 and 65, there was a high level of skill in the art of Factor V gene and protein when this application was filed. Specifically, both the Factor V protein sequence and the Factor V cDNA sequence were known in the art before the effective filing date of the present application. For example, Jenny *et al.* and Kane *et al.* reported complete cDNA and amino acid sequences of human Factor V in 1987. Jenny J.R. *et al.* is enclosed as Evidence Appendix C. Kane W.H. *et al.* is enclosed as Evidence Appendix D. In addition, Cripe *et al.* reported the genomic structure of the human Factor V gene in 1992. Cripe L.D. *et al.* is enclosed as Evidence Appendix E.

As discussed above, all of the methods needed to conduct nucleic acid sequencing assays were well known in the art when the application was filed. In the art of molecular biology when this application was filed, it was routine for one of ordinary skill in the art to isolate nucleic acids from a cell sample, to conduct a nucleic acid sequencing assay to determine a gene sequence and to compare the determined sequence to a known sequence in order to determine a mutation. For example, general methods and tools were described in Alberts B. *et al.* (eds.), 1986, "Molecular

Biology of the Cell," Garland Publishing, Inc., New York & London, pp 185-196 (Evidence Appendix F). Since the publication of Alberts B. *et al.*, methods of sequencing had been further developed and became one of the most essential skills of ordinary artisan in the field of molecular biology at the time when the present application was filed. Methods and tools (*e.g.*, primers and other reagents) specific for the isolation, amplification and sequencing of the Factor V cDNA directly from human tissues, *e.g.*, lymphocytes, were disclosed in Shen *et al.*, April 1, 1993, "The Serine Protease Cofactor V Is Synthesized by Lymphocytes," J. Immunology, 150:2992-3001 (Evidence Appendix G).

Furthermore, it was well within routine skills of an ordinary artisan to determine what sequence substitution would be silent or cause only conservative substitution in the Factor V gene and what sequence substitution would alter Factor V's function. For example, similar or conservative amino acids in general were described in Dayoff *et al.*, 1978, Atlas of Protein Sequence and Structure, M. O. Dayoff, ed., Nat'l Biomed. Research Fnd., Washington D.C., vol. 5, Suppl. 3, pp. 345-362. With respect to Factor V, Shen *et al.* discussed in detail that there are certain mutations in the Factor V gene that are silent or only cause conservative substitutions and may not significantly or only subtly affect the Factor V protein function. See, page 2996, left column of Evidence Appendix G.

Thus, one of ordinary skill in the art would readily have been able to isolate and amplify nucleic acid from the lymphocytes of an individual who had APC-resistance phenotype or was at risk of APC-resistance using specific methods and reagents described in Shen *et al.* (Evidence Appendix G). One of skill would also readily have been able to sequence the nucleic acid using routine sequencing methods known in the art and reagents specific for the Factor V gene described in Shen *et al.* (Evidence Appendix G). One of skill would readily have been able to determine a presence of an abnormal sequence in the individual's Factor V gene simply by comparing the determined sequence to a normal Factor V sequence disclosed by Jenny *et al.* (Evidence Appendix C) and Kane *et al.* (Evidence Appendix D). Then, one of skill would readily have been able to determine whether a particular abnormal sequence was a mutation associated with APC-resistance or was a silent mutation causing only conservative substitutions based on the common knowledge in the art and the teachings of Shen *et al.* (Evidence Appendix G).

As discussed above, the Federal Circuit stated in the Invitrogen decision that the purpose of the enablement requirement under section 112 requires that the patent specification enable those skilled in the art to make and use the full scope of the claimed invention without "undue experimentation" in order to extract meaningful disclosure of the invention and, by this disclosure, advance the technical arts. Invitrogen, 429 F.3d 1052. Appellant submits that the disclosure of the present specification precisely satisfies such an enablement requirement. In fact, a particular mutation in the Factor V gene that caused APC-resistance was determined by a skilled artisan, based upon the disclosure of the present application armed with the knowledge in the art, *e.g.*, the sequence of Factor V gene and methods of sequencing, without undue experimentation.

On June 18, 1994, Voorberg *et al.* reported that a mutation at position Arg506 in the Factor V protein was responsible for thromboembolism associated with APC-resistance based on the disclosure of the present application and the knowledge available in the art. J. Voorberg *et al.*, June 18, 1994, "Association of idiopathic venous with single point mutation at Arg506 of factor V," The Lancet, 343:1535-1536, a copy of which is enclosed as Evidence Appendix H. Evidence Appendix H was submitted together with a response to an Office Action on September 28, 2005, and was entered in the record by the Examiner. Voorberg *et al.* cited as reference 7, Dahlbäck B. *et al.*, February, 1994, "Inherited resistance to activated protein C is corrected by anticoagulant cofactor activity found to be a property of factor V," PNAS USA, 91:1396-1400, which summarizes the core discovery of the present application and was published in February, 1994, soon after the present application was filed.

Specifically, Voorberg *et al.* analyzed 27 patients with documented idiopathic thromboembolism by APC-resistant test disclosed in the present application (see, Evidence Appendix H, page 1535, left column, paragraph 3). Then, Voorberg *et al.* isolated RNA from peripheral blood lymphocytes as described by Shen *et al.* (Evidence Appendix G) and prepared cDNA using techniques known in the art at the time of invention. The cDNA was amplified by standard PCR and sequenced using standard methods known in the art. The sequence was compared with the known cDNA sequence of a normal Factor V gene. The comparison showed that the patients having APC-resistance contained an abnormal sequence at position 506, a Gln instead of a normal Arg. Thus, Voorberg *et al.* determined a specific mutation in the Factor V

gene determined to be associated with APC-resistance based on the teachings and disclosure of the present application and the knowledge known in the art at the time of invention without undue experimentation.

Therefore, the present specification fully complies with the enablement requirement with respect to claim 54 because it enabled those skilled in the art (*e.g.*, Voorberg) to make and use the full scope of the claimed invention without undue experimentation in order to extract meaningful disclosure of the invention and, by this disclosure, advance the technical arts.

For at least the above reasons, Appellant respectfully submit that the present application fully complies with the enablement requirement with respect to independent claim 54 and its dependent claim 55 and request reconsideration and withdrawal of the rejection of claims 54 and 55 under 35 U.S.C. § 112, first paragraph.

Respectfully submitted,

Date: June 14, 2006
Reg. No. 51,551

Fangli Chen, Ph.D.
Agent for the Appellant
Kirkpatrick & Lockhart Nicholson
Graham LLP
State Street Financial Center
One Lincoln Street
Boston, Massachusetts 02111-2950

Tel. No.: (617) 261-3198
Fax No.: (617) 261-3175

961697

CLAIMS APPENDIX

1-45. (Canceled)

46. (Previously presented) A method for detecting an individual at risk of developing thrombosis, said method comprising:

(a) obtaining a sample from the individual;

(b) conducting a nucleic acid assay on the sample, wherein the nucleic acid assay is a hybridization assay or a sequencing assay;

(c) determining abnormal presence or absence of at least one nucleic acid fragment or sequence in the individual's Factor V gene compared to a normal control; and

(d) detecting the individual at risk of developing thrombosis based on the determination of step (c).

47-52. (Canceled)

53. (Previously presented) The method of claim 46, wherein the nucleic acid assay is a sequencing assay.

54. (Previously presented) A method for determining a presence of a Factor V gene mutation associated with Activated Protein C (APC)-resistance in an individual at risk for APC-resistance, the method comprising the steps of:

(a) obtaining a sample from the individual;

(b) conducting a nucleic acid sequencing assay on the sample using reagents specific for the Factor V gene to determine the Factor V gene sequence; and

(c) determining the presence of the Factor V gene mutation associated with APC-resistance in the individual by comparing the sequence of the Factor V gene from step (b) to a normal Factor V gene sequence.

55. (Previously presented) The method of claim 54, wherein the mutation is determined as an abnormal absence or presence of at least one nucleotide sequence in the Factor V gene.

56-63. (Canceled)

64. (Previously presented) The method of claim 53, wherein the sequencing assay comprises sequencing the Factor V gene using reagents specific for the Factor V gene.

65. (Previously presented) The method of claim 64, wherein the detecting step detects an abnormal nucleotide sequence in the Factor V gene.

EVIDENCE APPENDIX

This appendix includes copies of the following evidence:

1. Evidence Appendix A relied upon by the Examiner in the Final Office Action as to the ground of rejection to be reviewed on appeal.

2. Evidence Appendix B relied upon by the Examiner in the Final Office Action as to the ground of rejection to be reviewed on appeal.

3. Evidence Appendix C submitted together with the Amendment and Response filed on September 28, 2005, and entered in the record by the Examiner.

4. Evidence Appendix D submitted together with the Amendment and Response filed on September 28, 2005, and entered in the record by the Examiner.

5. Evidence Appendix E submitted together with the Amendment and Response filed on September 28, 2005, and entered in the record by the Examiner.

6. Evidence Appendix F submitted together with the Amendment and Response filed on September 28, 2005, and entered in the record by the Examiner.

7. Evidence Appendix G submitted together with the Amendment and Response filed on September 28, 2005, and entered in the record by the Examiner.

8. Evidence Appendix H submitted together with a response to an Office Action on September 28, 2005, and entered in the record by the Examiner.

Evidence Appendix A

GENOME RESEARCH

A Closer Look at SNPs Suggests Difficulties

Using the wildly popular genome markers called SNPs to track genes may be less straightforward than researchers expected

SKOKLOSTER, SWEDEN—During the past year, single-nucleotide polymorphisms, commonly referred to as SNPs (pronounced snips), have taken the genomics community by storm. SNPs are single-base variations in the genetic code that occur about once every 1000 bases along the 3-billion-base human genome. Many researchers think that knowing the locations of these closely spaced DNA landmarks will ease both the sequencing of the human genome and the discovery of genes involved in such major human diseases as asthma, diabetes, atherosclerosis, schizophrenia, and cancer. But earlier this month at the first international meeting devoted to SNPs,* enthusiasts heard sobering news.

Although no one doubts that SNPs will ultimately prove to have some value in tracking disease genes and understanding human genetic diversity, new results presented at the meeting suggest that the task could prove more difficult than many had initially thought. In some cases, SNPs might fail to pick up disease genes, or researchers will need to have many more SNPs located in and around the suspected disease gene to make their case than first anticipated. Other work suggests researchers will also need more information about the history of the people being studied, such as their migration patterns, to make sense of their SNP data.

By the end of the 3-day conference, even the organizers went home shaking their heads. "There are so many problems and unanswered questions," complained Anthony Brookes, a co-organizer and geneticist from the University of Uppsala in Sweden. "At the moment, we're finding our way in the dark."

SNPs seem simple in part because the wealth of genome data being generated by the Human Genome Project and a range of faster, cheaper ways to find SNPs (*Science*, 15 May, p. 1077; 17 July, p. 363) are causing these markers to pile up quickly in both public and private databases. They are much more plentiful than other markers, such as microsatellites, used as genetic landmarks for tracking genes. And they have the added advantage of existing within genes as well as near them, possibly making them useful in identifying the specific variant of the gene that causes disease.

Indeed, most previous gene hunts required studying large, multigenerational families. But in 1996, epidemiologists Neil Risch at Stanford University in California and Kathleen Merikangas at Yale University in New Haven, Connecticut, suggested that SNPs might even be used to track down genes in unrelated people, particularly when the gene merely increases the risk for a disease. This would involve looking for differences in the patterns of SNPs between healthy and unhealthy people (*Science*, 13 September 1996, p. 1516). Prospects such as those led prominent geneticists, such as Francis Collins, director of the National Human Genome Research Institute, and Aravinda Chakravarti of Case Western Reserve University in Cleveland, Ohio, to propose that researchers find enough SNPs to perform such association studies (*Science*, 28 November 1997, p. 1580).

GRAPHIC OMITTED PER PUBLISHER UMI

Missed gene. SNPs didn't help Rosalind Harding find the sickle cell gene.

But as two groups reported at the Skokloster meeting, using SNPs to track genes may be less straightforward than thought. Both groups had problems in trying to use patterns of DNA variation to link test genes to diseases with which they were already known to be associated—heart disease in one case and sickle cell anemia in the other.

Working with Charles Sing from the University of Michigan, Ann Arbor, and his colleagues, population geneticist Andrew Clark from Pennsylvania State University in University Park focused on heart disease risk, first examining the role of the lipoprotein lipase (*LPL*) gene. Previous studies had shown that this gene, when mutated, causes high blood lipid concentrations and an increased incidence of heart disease in some families. Clark and his colleagues decided to use SNPs to find out which, if any, *LPL* gene variants might be increasing the risk for heart disease in the general population.

To do this, Sing's team first sequenced a 9700–base pair region of DNA containing the *LPL* gene in samples obtained from 24 people from each of three populations: one in Finland, the second in Rochester, Minnesota, and the third in Jackson, Mississippi. The researchers found that the region contained 88 SNPs, seven of which were in the protein-coding regions of the gene.

Clark and colleagues wanted to use the SNPs in epidemiological studies aimed at understanding the complex chain of genetic and environmental factors that affect heart disease risk. To do this, they first tried to construct a tree representing the historical sequence of mutations that gave rise to the SNPs. The idea was to group different variants of the *LPL* gene according to their ancestral relationships and then compare disease risk among the different lineages. But it immediately became clear that this would be difficult if not impossible; parts of the gene had been shuffled by recombination, the DNA exchanges that occur between the maternal and paternal copies of a gene during sperm and egg formation.

GRAPHIC OMITTED PER PUBLISHER UMI

SNP hunt hurdle. Joseph Terwilliger cites perils of genome complexity.

"There has been almost as much recombination as mutation," Clark reported, and that, he adds, "is going to make SNP mapping and association tests much more difficult." Recombination can break down the correlation between the SNPs and variants that inflate disease risk, making it much harder to identify the association. Everybody hopes that only a few regions of the human genome will exhibit this high level of recombination, but where it does occur researchers will need many more SNPs to increase the odds of finding some that correlate with the pertinent mutations.

Rosalind Harding of the John Radcliffe Hospital in Oxford, United Kingdom, shares Clark's concerns about the utility of SNPs, particularly if researchers try to depend on SNPs alone to identify disease genes. A single-base change in the β-globin gene has long been known to cause sickle cell anemia, and she tried to see if SNPs would re-

* The 1st International Meeting on Single-Nucleotide Polymorphism and Complex Genome Analysis was held in Skokloster, Sweden, 29 August to 1 September.

More SNPs on the Way

Late last year, the National Cancer Institute (NCI) launched a project to find genome markers called single-nucleotide polymorphisms, or SNPs, to use in tracking down the hundreds of genes thought to affect cancer risk. NCI has already put about $1 million into the project, called the Genetic Annotation Initiative (GAI), which began generating SNPs in the spring. Researchers running the initiative are hoping that their approach will avoid many of the problems in using SNPs discussed at a recent conference in Skokloster, Sweden (see main text).

NCI is taking what NCI geneticist Ken Buetow, who oversees the GAI project, calls a "gene-based" approach. Instead of creating a genomewide map of anonymous SNPs, Buetow says, NCI will look for SNPs in the coding regions, and in the sequences at both ends, of several thousand genes suspected of contributing to cancer susceptibility or resistance. Besides the 100-plus known cancer-promoting oncogenes and the three dozen or so tumor suppressor genes, the pool will include DNA repair genes, genes that drive the cell division cycle, and genes involved in drug metabolism, immune responses, embryonic development, and cell migration and metastasis. Genes from the NCI's huge Cancer Genome Anatomy Project, which aims for a complete genetic profile of cancer cells (*Science*, 16 May 1997, p. 1023), will also be included as they're identified.

Buetow expects the average gene to yield three to five SNPs, a marker density that makes it much more likely that at least one will be close enough to any cancer mutation to be inherited with it as a block—a phenomenon called linkage disequilibrium. That doesn't ensure researchers won't miss the mutation when screening cancer patients—one of the researchers describing SNP problems at Skokloster had just such an experience with the sickle cell gene—but it should help. "We are less dependent on linkage disequilibrium relationships existing over long distances," says Buetow. "We're going to be right inside the genes." The data generated by GAI will also help determine how common the problems reported at the meeting are.

Once identified, the SNPs will be posted on a database of the National Center for Biotechnology Information, where researchers can access them and design and conduct "association studies" to see if the SNP patterns of cancer patients are different from those of controls. The hoped-for result: hundreds of new cancer genes. Cancer researchers welcome the new initiative. "Given the present technology, it seems to be the obvious next step," says Sofia Merajver, a breast cancer researcher at the University of Michigan, Ann Arbor.

Several issues are still up in the air, however. NCI hasn't decided which populations to screen for SNPs. Right now it's using the DNA of four people from the largely Caucasian families collected at the Centre d'Études du Polymorphisme Humain in France. The GAI wants more diversity, but no one agrees on what that means. "There is concern about stigmatization of populations and concern about what is a representative population," says Buetow. "There are going to be dramatic differences [in SNP frequency] based on geography."

Also under debate is the question of how deep to dig for cancer SNPs. Some would be satisfied with the common ones, in which case screening as few as eight individuals should yield the vast majority. But others argue that the newer, rarer SNPs are also needed, because they're more often in linkage disequilibrium with cancer mutations and thus more likely to show up in cancer association studies.

But the biggest question mark is what technology will be used to discover SNPs and then to detect or "score" them in cancer patients. "Not only has this not been done on a mass scale, but new technologies are being developed so fast, it's hard to know what to do," says the NCI's Mike Dean. To begin, Dean is using a high-performance liquid chromatography mutation-detection method developed by Peter Oefner of Stanford University. Buetow is doing conventional gel-based sequencing, which would be tedious and expensive for large-scale studies.

GRAPHIC OMITTED PER PUBLISHER UMI

Gene guide? SNPs, such as the cytosine (C) to thymine (T) change shown here, may point to cancer genes.

One technology now in high demand is the DNA "chip," which can quickly identify SNPs across long stretches of DNA. Affymetrix, a Santa Clara, California, biotech company, has developed such chips, which researchers at the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology and Affymetrix are using to do SNP prospecting (*Science*, 15 May, p. 1077). The National Institutes of Health is now negotiating with Affymetrix for a license, and both parties are optimistic. "We would be very happy to collaborate with the NIH in the area of SNP discovery," says Robert Lipshutz, Affymetrix's vice president of corporate development.

Whatever the outcome, Buetow is optimistic about finding methods that will make all kinds of cancer gene discovery projects easy. "We hope to push the technology to enable investigators to do any kind of study they want to do," he says. —KEN GARBER

Ken Garber is a science and health writer in Ann Arbor, Michigan.

veal the mutant gene. By analyzing DNA samples from 500 people randomly selected from around the world, Harding and her colleagues found that the β-globin gene has dozens of SNPs located in and around its coding sequences.

One of these SNPs turned out to be the sickle cell mutation itself. But when Harding looked at the frequencies of individual SNPs in the 500 samples, and also at inherited SNP patterns called haplotypes, searching for some sign that a particular SNP or haplotype was different, she found nothing that pointed to the sickle cell mutation. With SNP data alone, Harding concluded, "there will be nowhere near enough information to find something unusual and say 'there's a disease gene.'"

She predicts that, in addition to relying on SNPs, researchers will need to know about the patterns of disease and the history of the people being studied. "There has been this naïve idea that once you've gotten to the gene, you'll be able to decide which is the [pertinent] mutation," she adds. "But this is going to be very hard." Others concur. "You can't have just SNPs on their own," says Nigel Spurr, a geneticist with SmithKline Beecham in Harlow, United Kingdom. "You must have [other information and technology] to go with it."

For statistical geneticist Joseph Terwilliger of Columbia University in New York City, Harding's and Clark's experiences with SNPs are indicative of the underappreciated complexity of the genome and of the pitfalls of thinking SNPs will easily

lead geneticists to elusive disease genes. "Risch and Merikangas have been taken out of context" by overly enthusiastic promoters of SNPs' potential, he argues.

Terwilliger notes that although Risch and Merikangas found association studies practical for identifying disease genes in which one mutation accounts for most of the increased risk, that situation may be uncommon. In a survey of all the new disease genes reported in the *American Journal of Human Genetics* during the past 1.5 years, Terwilliger found that about 90% of those genes had more than 10 pertinent mutations that predispose an individual to disease. With so many different mutations involved, none is likely to stand out in a SNP analysis. And that's the easy case, involving diseases caused by mutations in a single gene. The situation will be worse for cancer and the many other diseases in which multiple genes contribute to increased risk. "It's not just the underestimated complexity of the genome as much as it is the underestimated complexity of the etiology of a complex disease," he adds.

GENSET

GRAPHIC OMITTED PER PUBLISHER UMI

Optimist. Genset's Daniel Cohen says SNP problems can be solved.

Researchers at Skokloster agreed that it will be difficult to gauge the usefulness of SNPs until they know more about how genomes vary between and within the world's ethnic groups. Because the most universal SNPs will be among the oldest, they are likely to exist in people both with and without disease. This means that there may be no distinctive pattern of SNPs specifically associated with a key variant of a gene. It could be easy to miss an important association or to make an association with the wrong gene variant. "If we don't think carefully before we do these experiments, we'll wind up with a lot of false signals," Uppsala's Brookes says.

Others at the meeting pointed out that association studies require that researchers look at much larger numbers of people than typical family studies, to sift out the false signals. "It's not enough to have 70 controls and 50 patients," says Gert-Jan Van Ommen, head of the Human Genome Organization and a geneticist at Sylvius Laboratories in Lieden, the Netherlands. "You're talking about requiring populations of several thousand." SNP analysis won't begin to be useful without new, high-speed technology for analyzing the thousands of DNA samples required, says Spurr.

Even with these caveats, however, the researchers expect to see SNPs research proceed. "We know they will be successful in certain situations," comments Case Western's Chakravarti. "We just don't know how successful they will be."

Already, association studies have linked a few gene variants to diseases. The tying of the *ApoE4* gene to an increased risk for Alzheimer's disease in Caucasians is one often-cited example. And geneticist Daniel Cohen, head of Genset in Evry, France, says that his company has worked out many of the issues raised by the conference participants, in part by developing new methods—which he would not describe in detail—for analyzing the data and discerning real associations. In October, for example, he plans to announce the SNP-based discovery of two genes involved in prostate cancer. "I am absolutely confident of this strategy," says Cohen. "It works."

Although others may not share Cohen's confidence, they want SNPs to be put to work. "Provided they are not being regarded as the panacea for complex disease findings, there is value in producing SNPs," says Van Ommen. "[SNPs] are going to make a big difference." —ELIZABETH PENNISI

ECOLOGY

Software Helps Australia Manage Forest Debate

A computer program to promote biodiversity gives loggers and conservationists a chance to end their fierce fighting over forest reserves

MELBOURNE, AUSTRALIA—The forests of New South Wales (NSW) have seen many bitter battles in the last 20 years between logging interests eager to feed an insatiable Japanese appetite for wood pulp and conservationists trying to preserve the country's dwindling arboreal heritage. Those battles have taken a heavy toll on the participants. Just ask Col Dorber, the executive director of NSW Forest Products Association. In 1995, Dorber suffered a stress-induced heart attack after being roundly condemned by government and industry officials and vilified in the media for publicly defending a logger caught punching a "greenie."

TOM BARRETT

Now back on the job, Dorber sees his remarks as an unfortunate reflection of the historic enmity between the two camps. That's why he's so encouraged by an experiment drawing attention from ecologists and resource managers around the world that attempts to inject science into forest management and that respects the interests of all parties. "Since 1995, we've been through a culture change," he says. "Prior to that, we [industry and conservationists] wouldn't speak to each other. But now we've learnt to respect each other. It's a fantastic process."

That process is a joint initiative by the federal and state governments to negotiate long-term agreements for forest reserves that allow continued logging while maximizing biodiversity. At the core of the negotiations is a computer program, called C-Plan, that gives adversaries a chance to trade in their swords for software. Like some ecological card game, the software puts a biodiversity value on each parcel of land and presents stakeholders with various packages that meet the conservation target. C-Plan was developed by NSW National Parks and Wildlife Service conservation planners Bob Pressey, Simon Ferrier, and colleagues, and programmer Mathew Watts at the University of New England in Armidale, NSW. So far it has been used in two major sets of negotiations; a third exercise, involving a large swathe of old-growth forest, has just begun.

"It's setting the gold standard in the field," says ecologist Reed Noss, co-executive director of the Conservation Biology Institute in Corvallis, Oregon, and president of the international Society for Conservation Biology. Indeed, the World Bank is using C-Plan for an assessment in Guyana, and Pressey is currently in South Africa to help

GRAPHIC OMITTED PER PUBLISHER UMI

Green software. Negotiators use C-Plan to help select reserve areas in Eden forest.

lead geneticists to elusive disease genes. "Risch and Merikangas have been taken out of context" by overly enthusiastic promoters of SNPs' potential, he argues.

Terwilliger notes that although Risch and Merikangas found association studies practical for identifying disease genes in which one mutation accounts for most of the increased risk, that situation may be uncommon. In a survey of all the new disease genes reported in the *American Journal of Human Genetics* during the past 1.5 years, Terwilliger found that about 90% of those genes had more than 10 pertinent mutations that predispose an individual to disease. With so many different mutations involved, none is likely to stand out in a SNP analysis. And that's the easy case, involving diseases caused by mutations in a single gene. The situation will be worse for cancer and the many other diseases in which multiple genes contribute to increased risk. "It's not just the underestimated complexity of the genome as much as it is the underestimated complexity of the etiology of a complex disease," he adds.

GENSET

GRAPHIC OMITTED PER PUBLISHER UMI

Optimist. Genset's Daniel Cohen says SNP problems can be solved.

Researchers at Skokloster agreed that it will be difficult to gauge the usefulness of SNPs until they know more about how genomes vary between and within the world's ethnic groups. Because the most universal SNPs will be among the oldest, they are likely to exist in people both with and without disease. This means that there may be no distinctive pattern of SNPs specifically associated with a key variant of a gene. It could be easy to miss an important association or to make an association with the wrong gene variant. "If we don't think carefully before we do these experiments, we'll wind up with a lot of false signals," Uppsala's Brookes says.

Others at the meeting pointed out that association studies require that researchers look at much larger numbers of people than typical family studies, to sift out the false signals. "It's not enough to have 70 controls and 50 patients," says Gert-Jan Van Ommen, head of the Human Genome Organization and a geneticist at Sylvius Laboratories in Lieden, the Netherlands. "You're talking about requiring populations of several thousand." SNP analysis won't begin to be useful without new, high-speed technology for analyzing the thousands of DNA samples required, says Spurr.

Even with these caveats, however, the researchers expect to see SNPs research proceed. "We know they will be successful in certain situations," comments Case Western's Chakravarti. "We just don't know how successful they will be."

Already, association studies have linked a few gene variants to diseases. The tying of the *ApoE4* gene to an increased risk for Alzheimer's disease in Caucasians is one often-cited example. And geneticist Daniel Cohen, head of Genset in Evry, France, says that his company has worked out many of the issues raised by the conference participants, in part by developing new methods—which he would not describe in detail—for analyzing the data and discerning real associations. In October, for example, he plans to announce the SNP-based discovery of two genes involved in prostate cancer. "I am absolutely confident of this strategy," says Cohen. "It works."

Although others may not share Cohen's confidence, they want SNPs to be put to work. "Provided they are not being regarded as the panacea for complex disease findings, there is value in producing SNPs," says Van Ommen. "[SNPs] are going to make a big difference." —ELIZABETH PENNISI

ECOLOGY

Software Helps Australia Manage Forest Debate

A computer program to promote biodiversity gives loggers and conservationists a chance to end their fierce fighting over forest reserves

MELBOURNE, AUSTRALIA—The forests of New South Wales (NSW) have seen many bitter battles in the last 20 years between logging interests eager to feed an insatiable Japanese appetite for wood pulp and conservationists trying to preserve the country's dwindling arboreal heritage. Those battles have taken a heavy toll on the participants. Just ask Col Dorber, the executive director of NSW Forest Products Association. In 1995, Dorber suffered a stress-induced heart attack after being roundly condemned by government and industry officials and vilified in the media for publicly defending a logger caught punching a "greenie."

TOM BARRETT

Now back on the job, Dorber sees his remarks as an unfortunate reflection of the historic enmity between the two camps. That's why he's so encouraged by an experiment drawing attention from ecologists and resource managers around the world that attempts to inject science into forest management and that respects the interests of all parties. "Since 1995, we've been through a culture change," he says. "Prior to that, we [industry and conservationists] wouldn't speak to each other. But now we've learnt to respect each other. It's a fantastic process."

That process is a joint initiative by the federal and state governments to negotiate long-term agreements for forest reserves that allow continued logging while maximizing biodiversity. At the core of the negotiations is a computer program, called C-Plan, that gives adversaries a chance to trade in their swords for software. Like some ecological card game, the software puts a biodiversity value on each parcel of land and presents stakeholders with various packages that meet the conservation target. C-Plan was developed by NSW National Parks and Wildlife Service conservation planners Bob Pressey, Simon Ferrier, and colleagues, and programmer Mathew Watts at the University of New England in Armidale, NSW. So far it has been used in two major sets of negotiations; a third exercise, involving a large swathe of old-growth forest, has just begun.

"It's setting the gold standard in the field," says ecologist Reed Noss, co-executive director of the Conservation Biology Institute in Corvallis, Oregon, and president of the international Society for Conservation Biology. Indeed, the World Bank is using C-Plan for an assessment in Guyana, and Pressey is currently in South Africa to help

GRAPHIC OMITTED PER PUBLISHER UMI

Green software. Negotiators use C-Plan to help select reserve areas in Eden forest.

Evidence Appendix B

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 14/47, C12N 15/12, C12Q 1/68, C07K 16/18, A01K 67/027, G06F 17/30, 17/50

A2

(11) International Publication Number: **WO 99/52942**

(43) International Publication Date: 21 October 1999 (21.10.99)

Express Mail EV832482014US
Mailing Label No.

(21) International Application Number: PCT/IB99/00744

(22) International Filing Date: 15 April 1999 (15.04.99)

(30) Priority Data:

60/081,893	15 April 1998 (15.04.98)	US
60/091,314	30 June 1998 (30.06.98)	US
60/123,406	8 March 1999 (08.03.99)	US

(71) Applicant *(for all designated States except US)*: GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). CHUMAKOV, Ilya [FR/FR]; 196, rue des Chèvrefeuilles, F-77000 Vaux-le-Penil (FR). BOUGUELERET, Lydie [FR/FR]; 14, rue Vouillé, F-75015 Paris (FR).

(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: GENOMIC SEQUENCE OF THE 5-LIPOXYGENASE-ACTIVATING PROTEIN (FLAP), POLYMORPHIC MARKERS THEREOF AND METHODS FOR DETECTION OF ASTHMA

(57) Abstract

The invention concerns the genomic sequence of the *FLAP* gene. The invention also concerns biallelic markers of a *FLAP* gene and the association established between these markers and diseases involving the leukotriene pathway such as asthma. The invention provides means to determine the predisposition of individuals to diseases involving the leukotriene pathway as well as means for the diagnosis of such diseases and for the prognosis/detection of an eventual treatment response to agents acting on the leukotriene pathway.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Genomic Sequence Of The 5-Lipoxygenase-Activating Protein (FLAP), Polymorphic Markers Thereof And Methods For Detection Of Asthma.

FIELD OF THE INVENTION

The invention concerns the genomic sequence of the *FLAP* gene. The invention also concerns biallelic markers of a *FLAP* gene and the association established between these markers and diseases involving the leukotriene pathway such as asthma. The invention provides means to determine the predisposition of individuals to diseases involving the leukotriene pathway as well as means for the diagnosis of such diseases and for the prognosis/detection of an eventual treatment response to agents acting on the leukotriene pathway.

BACKGROUND OF THE INVENTION

The progression of inflammatory diseases in which the synthesis of leukotrienes plays an active role, such as asthma and arthritis, constitutes a major health problem in Western societies.

For example, the prevalence of asthma in Occidental countries has risen steadily over the last century, affecting about 10% of the population. In 1994, it afflicted more than 14 million people in the United States alone (including 4.8 million (6.9%) less than 18 year of age) whereas only 8 million people suffered from the same disease in 1982. It claims more than 5000 lives each year (including 342 deaths among persons aged less than 25 in 1993). Asthma affects one child in seven in Great Britain, and in the United States, it causes one-third of pediatric emergency-room visits. It is the most frequent chronic disease in childhood.

Bronchial asthma is a multifactorial syndrome rather than a single disease, defined as airway obstruction and characterized by inflammatory changes in the airways and bronchial hyper-responsiveness. Stimuli which cause the actual asthma attacks include allergens (in sensitized individuals), exercise (in which one stimulus may be cold air), respiratory infections and atmospheric pollutants such as sulphur dioxide. The asthmatic subject has intermittent attacks of dyspnoea (difficulty in breathing out), wheezing, and cough that can be life-threatening or even fatal.

The manifestation of asthma probably involves both genetic and environmental factors, and in most subjects the asthmatic attack consists of two phases which illustrate the pathophysiology of the condition:

- an immediate phase, consisting mainly of bronchospasms due to spasms of the bronchial smooth muscle; the cells involved are mast cells releasing histamine, but also eosinophils, macrophages and platelets releasing leukotrienes, prostaglandins, and platelet-activating factor; these spasmogens added to chemotaxins and chemokins attract leukocytes into the area, setting the stage for the delayed phase;
- a later phase consisting of a special type of inflammation comprising vasodilatation, oedema, mucus secretion and bronchospasm; it is caused by inflammatory mediators released from

activated cytokine-releasing T cells and eosinophils, and, possibly, neuropeptides released by axon reflexes; these mediators cause damage and loss of bronchial epithelium.

The strongest identifiable predisposing factor for developing asthma is atopy, the predisposition for the development of an IgE-mediated response to common aeroallergens. When IgE binds to the IgE receptors on the cells, the system becomes primed so that subsequent re-exposure to the relevant allergen will cause an asthmatic attack. Most asthma cases (95%) are associated with atopy.

Further to their above-mentioned role in asthma, leukotrienes are more generally involved in host defense reactions and play an important role in immediate hypersensitivity as well as in inflammatory diseases other than asthma such as inflammatory bowel disease, psoriasis and arthritis.

The leukotriene pathway

Leukotrienes are products of the Lipoxygenase pathways. Lipoxygenases are soluble enzymes located in the cytosol and are found in lung, platelets, mast cells, and white blood cells. The main enzyme in this group is 5-Lipoxygenase which is the first enzyme in the biosynthesis of leukotrienes.

The first step in leukotriene biosynthesis is the release of arachidonic acid from membrane phospholipids upon cell stimulation (for example, by immune complexes and calcium ionophores). Arachidonic acid is then converted into leukotrienes A4 by a 5-Lipoxygenase (5-LO) which translocates to the cell membrane where it becomes associated to a protein called "five-Lipoxygenase activating protein" (FLAP), which is necessary for leukotriene synthesis in intact cells. 5-LO also has leukotriene A4 hydrolase activity.

Leukotriene A4 (LTA4), an unstable epoxide intermediate, is then hydrolyzed into leukotriene B4 (LTA4-hydrolase activity) or conjugated with glutathione to yield leukotriene C4 (LTC4-synthase activity) and its metabolites, leukotriene D4 and leukotriene E4. LTB4 is produced mainly by neutrophils, while cysteinyl-leukotrienes (LTC4, LTD4, and LTE4) are mainly produced by eosinophils, mast cells, basophils, and macrophages.

LTB4 is a powerful chemotactic agent for both neutrophils and macrophages. On neutrophils, it also causes up-regulation of membrane adhesion molecules and increases the production of toxic oxygen products and the release of granule enzymes. On macrophages and lymphocytes, it stimulates proliferation and cytokine release. Thus LTB4 is an important mediator in all types of inflammations.

Cysteinyl-leukotrienes act on the respiratory and cardiovascular systems. In the respiratory system, they are potent spasmogens causing a contraction of bronchiolar muscle and an increase in mucus secretion. In the cardiovascular system, they cause vasodilatation in most vessels, but they also act as coronary vasoconstrictors. The cysteinyl-leukotrienes are of particular importance in asthma.

FLAP (5-lipoxygenase-activating protein)

FLAP is a 18-kD membrane-bound polypeptide which specifically binds arachidonic acid and activates 5-LO by acting as an arachidonic acid transfer protein. The *FLAP* gene spans greater than 31 kb and consists of five small exons and four large exons (See GenBank 182657, Kennedy et al. 1991 incorporated herein by reference, Genbank M60470 for exon 1, Genbank M63259 for exon 2, Genbank M63260 for exon 3, Genbank M63261 for exon 4, and Genbank M6322 for exon 5).

The nuclear envelope is the intracellular site at which 5-LO and FLAP act to metabolize arachidonic acid, and ionophore activation of neutrophils and monocytes results in the translocation of 5-LO from a nonsedimentable location to the nuclear envelope. Inhibitors of FLAP function prevent translocation of 5-LO from cytosol to the membrane and inhibit 5-LO activation. They are thus interesting anti-inflammatory drug candidates. Indeed, antagonists of FLAP can attenuate allergen-induced bronchoconstrictor responses which supports an important role for cysteinyl leukotrienes in mediating these asthmatic responses.

Pharmacogenomics

To assess the origins of individual variations in disease susceptibility or drug response, pharmacogenomics uses the genomic technologies to identify polymorphisms within genes that are part of biological pathways involved in disease susceptibility, etiology, and development, or more specifically in drug response pathways responsible for a drug's efficacy, tolerance, or toxicity, including but not limited to drug metabolism cascades.

In this regard, the inflammatory phenomena which are involved in numerous diseases present a high relevance to pharmacogenomics both because they are at the core of many widespread serious diseases, and because targeting inflammation pathways to design new efficient drugs includes numerous risks of potentiating serious side-effects. The leukotriene pathway is particularly interesting since its products are powerful inflammatory molecules.

The vast majority of common diseases, such as cancer, hypertension and diabetes, are polygenic (involving several genes). In addition, these diseases are modulated by environmental factors such as pollutants, chemicals and diet. This is why many diseases are called multifactorial; they result from a synergistic combination of factors, both genetic and environmental.

For example, in addition to the evidenced impact of environmental factors on the development of asthma, patterns of clustering and segregation analyzes in asthmatic families have suggested a genetic component to asthma. However, the lack of a defined and specific asthma phenotype is proving to be a major hurdle for reliably detecting asthma-associated genes.

Asthma is usually diagnosed through clinical examination and biological testing. The non-specific bronchial hyper-responsiveness that accompanies asthma is measured by the variation of airflow triggered in a patient by the administration of a bronchoconstrictor such as histamine or methacholine. Atopy is detected by skin prick tests that measure serum IgE titers. Standard

symptom questionnaires are also commonly used to detect symptoms characteristic of, but not unique to, asthma (like nocturnal wheeze and breathlessness).

However, there is no straightforward physiological or biological blood test for the asthmatic state. Despite advances in understanding the pathophysiology of asthma and its development, evidence suggests that the prevalence of the asthmatic state and the severity of asthma attacks is underestimated. As a result, adequate asthma treatment is often delayed, thereby allowing the inflammation process to better establish itself. Thus, there is a need for an efficient and reliable asthma diagnostic test.

Drug efficacy and toxicity may also be considered as multifactorial traits that involve genetic components in much the same way as complex diseases. In this respect, there are three main categories of genes that may theoretically be expected to be associated with drug response, namely genes linked with the targeted disease, genes related to the drug's mode of action, and genes involved in the drug's metabolism.

The primary goal of pharmacogenomics in the study of asthma is to look for genes that are related to drug response. It can first provide tools to refine the design of drug development by decreasing the incidence of adverse events in drug tolerance studies, by better defining patient subpopulations of responders and non-responders in efficacy studies and, by combining the results obtained therefrom, to further allow for better individualized drug usage based on efficacy/tolerance prognosis.

Pharmacogenomics can also provide tools to identify new targets drug design and to optimize the use of already existing drugs, in order to either increase their response rate and/or exclude non-responders from particular treatments, or decrease undesirable side-effects and/or exclude from corresponding treatment patients with significant risk of undesirable side-effects.

For this second application of pharmacogenomics, the leukotrienes pathway is also useful because many anti-asthmatic and anti-inflammatory agents which act through the leukotrienes pathway are under development, most of which show some incidence of severe side-effects.

For example, there are two major categories of anti-asthma drugs: bronchodilators and anti-inflammatory agents. Bronchodilators are effective in reversing the bronchospasm of the immediate phase of the disease. Drugs used as bronchodilators include the β_2-adrenoceptor agonists (dilating the bronchi by a direct action on the smooth muscle, e.g. salbutamol), the xanthines (e.g. theophylline) and the muscarinic-receptor antagonists (e.g. ipratropium bromide). These represent the short term attack symptomatic treatment.

Anti-inflammatory agents are effective in inhibiting or preventing the production of inflammatory components in both asthma phases. They include glucocorticoids, sodium cromoglycate and histamine H1-receptor antagonists. These agents represent the current long term treatment of the asthmatic state.

However, none of these currently used anti-asthmatic drugs is completely satisfactory as none actually "cures" all patients with the disease. Glucocorticoids are the most interesting active compounds in this regard but they have potentially serious unwanted side-effects (oropharyngeal candidiasis, dysphonia and osteoporosis for inhaled glucocorticoids, and mood disturbances, increased appetite and loss of glucose control in diabetics for systemic glucocorticoids).

In recent years, more effective and selective leukotriene biosynthesis inhibitors (e.g., 5-LO and FLAP-binding inhibitors) have been developed and used as novel therapies for bronchial asthma and other inflammatory disorders. For example, Zileuton (Zyflo®), an inhibitor of 5-LO commercialized by Abbott Laboratories (Abbott Park, Illinois), has been shown to improve airway function and to reduce asthma-related symptoms.

Unfortunately, undesirable side-effects such as acute exacerbation of asthma, dyspepsia and elevated liver enzymes have been reported in clinical trials for Zileuton. There is also concern about drug interactions with hepatically cleared medicaments.

Thus, in addition to the need for the development of an efficient and reliable asthma diagnostic test, there is also a need to develop more effective and better targeted therapeutic strategies acting on the leukotrienes pathway with reduced side-effects and low toxicity for the user. One way to achieve this in the relative short term would be through the use of pharmacogenomics results, to better define the use of existing drugs or drug candidates in order to enhance the benefit/risk ratio on target subpopulations of patients.

SUMMARY OF THE INVENTION

The present invention stems from the isolation and characterization of the whole genomic sequence of the *FLAP* gene including its regulatory regions. Oligonucleotide probes and primers hybridizing specifically with a genomic sequence of *FLAP* are also part of the invention. A further object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described in the present invention, and in particular recombinant vectors comprising the regulatory region of *FLAP* or a sequence encoding the FLAP enzyme, as well as cell hosts comprising said nucleic acid sequences or recombinant vectors. The invention also encompasses methods of screening of molecules which modulate or inhibit the expression of the *FLAP* gene. The invention also comprises a new allelic variant of the FLAP protein.

The invention is also directed to biallelic markers that are located within the *FLAP* genomic sequence, these biallelic markers representing useful tools in order to identify a statistically significant association between specific alleles of *FLAP* gene and diseases involving the leukotriene pathway such as inflammatory diseases, or between specific alleles of *FLAP* gene and either side-effects resulting from the administration of agents acting on the leukotriene pathway, preferably Zileuton, or a beneficial response to treatment with agents acting on the leukotriene pathway. These associations are within the scope of the invention.

More particularly, the present invention stems from the identification of genetic associations between alleles of biallelic markers of the *FLAP* gene and asthma, as confirmed and characterized in a panel of human subjects.

Methods and products are provided for the molecular detection of a genetic susceptibility in humans to diseases involving the leukotriene pathway such as inflammatory diseases and comprising, among others, asthma, arthritis, psoriasis and inflammatory bowel disease. They can be used for diagnosis, staging, prognosis, and monitoring of such diseases, which processes can be further included within treatment approaches. The invention also provides for the efficient design and evaluation of suitable therapeutic solutions including individualized strategies for optimizing drug usage, and screening of potential new medicament candidates.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram of the *FLAP* gene with an indication of relative position of the biallelic markers of the present invention.

Figure 2 show the results of an association study between the *FLAP* biallelic markers and asthma with 290 asthmatic individuals and 280 US Caucasian controls. Figure 2 is a graph demonstrating the association between some of the biallelic markers of the invention and asthma with the absolute value of the logarithm (base 10) of the p-value of the chi-square values for each marker shown on the y-axis and a rough estimate of the position of each marker with respect to the *FLAP* gene elements on the x-axis.

Figure 3 is a table demonstrating the results of a haplotype association analysis between asthma and haplotypes which consist of biallelic markers of the invention. (297 cases vs 286 Caucasian US controls)

Figure 4 is a table demonstrating the results of a haplotype frequency analysis including permutation testing with more than 1000 iterations.

BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING

SEQ ID No 1 contains a genomic sequence of *FLAP* comprising the 5' regulatory region (upstream untranscribed region), the exons and introns, and the 3' regulatory region (downstream untranscribed region).

SEQ ID No 2 contains a complete human *FLAP* cDNA with 5' and 3' UTRs.

SEQ ID No 3 contains the FLAP protein encoded by the cDNA of SEQ ID No 2.

SEQ ID Nos 4 and 5 contain either allele 1 or 2 of the biallelic maker A14 and its surrounding sequence.

SEQ ID Nos 6 and 7 contain the sequence of amplification primers for the biallelic maker A14.

SEQ ID No 8 contains the sequence of a microsequencing primer of the biallelic maker A14.

SEQ ID Nos 9 and 10 contain either allele 1 or 2 of the biallelic maker A19 and its surrounding sequence.

SEQ ID Nos 11 and 12 contain the sequence of amplification primers for the biallelic maker A19.

SEQ ID No 13 contains the sequence of a microsequencing primer of the biallelic maker A19.

SEQ ID No 14 contains a primer containing the additional PU 5' sequence described further in Example 2.

SEQ ID No 15 contains a primer containing the additional RP 5' sequence described further in Example 2.

DETAILED DESCRIPTION OF THE INVENTION

5-LO is associated with FLAP for leukotriene synthesis. Indeed, it appears that regulation of the production of leukotrienes can be achieved either through the action of direct 5-LO inhibitors or indirect leukotriene biosynthesis inhibitors which bind to FLAP.

The present invention concerns the identification and characterization of biallelic markers in a FLAP encoding gene, as well as the identification of significant polymorphisms associated with diseases involving the leukotriene pathway. Preferably, the polymorphisms are associated with asthma.

The identified polymorphisms are used in the design of assays for the reliable detection of genetic susceptibility to diseases involving the leukotriene pathway. They can also be used in the design of drug screening protocols to provide an accurate and efficient evaluation of the therapeutic and side-effect potential of new or already existing medicaments.

I. Definitions

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

The term "*FLAP* gene", when used herein, encompasses genomic, mRNA and cDNA sequences encoding the FLAP protein. In the case of a genomic sequence, the *FLAP* gene also includes native regulatory regions which control the expression of the coding sequence of the *FLAP* gene.

As used interchangeably herein, the terms "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or

phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No WO 95/04064. However, the polynucleotides of the invention are preferably comprised of greater than 50% conventional deoxyribose nucleotides, and most preferably greater than 90% conventional deoxyribose nucleotides. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

The term "purified" is used herein to describe a polynucleotide or polynucleotide vector of the invention which has been separated from other compounds including, but not limited to other nucleic acids, carbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50 , preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50 , preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity may be indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "polypeptide" refers to a polymer of amino without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude prost-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from

mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "purified" is used herein to describe a polypeptide of the invention which has been separated from other compounds including, but not limited to nucleic acids, lipids, carbohydrates and other proteins. A polypeptide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure polypeptide typically comprises about 50%, preferably 60 to 90% weight/weight of a protein sample, more usually about 95%, and preferably is over about 99% pure. Polypeptide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen., which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', F(ab)$_2$, and F(ab')$_2$ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a FLAP polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by H. Mario Geysen et al. 1984; PCT Publication No WO 84/03564; and PCT Publication No WO 84/03506.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another be virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

The term "allele" is used herein to refer to variants of a nucleotide sequence. Diploid organisms may be homozygous or heterozygous for an allelic form.

A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell required to initiate the specific transcription of a gene.

A sequence which is "operably linked" to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the promoter to direct the transcription of the coding polynucleotide.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves as an initiation point for nucleotide polymerization catalyzed by either DNA polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined hereinbelow) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence complementary of the specific polynucleotide sequence to be identified.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population, which are heterozygous at a particular allele. In a biallelic system the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where P_a is the frequency of the least common allele. In order

to be useful in genetic studies a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which may be associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different genomes or between different individuals, the polymorphic site may be occupied by two different nucleotides.

"Biallelic markers" consist of a single base polymorphism. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence included in a gene, which, when compared with one another, present a nucleotide modification at one position. Usually, the nucleotide modification involves the substitution of one nucleotide for another (for example C instead of T). Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker."

As used herein the terminology "defining a biallelic marker" means that a sequence includes a polymorphic base from a biallelic marker. The sequences defining a biallelic marker may be of any length consistent with their intended use, provided that they contain a polymorphic base from a biallelic marker. The sequence has between 1 and 500 nucleotides in length, preferably between 5, 10 , 15, 20, 25, or 40 and 200 nucleotides and more preferably between 30 and 50 nucleotides in

length. Preferably, the sequences defining a biallelic marker include a polymorphic base selected from the group consisting of biallelic markers A1 to A28. In some embodiments the sequences defining a biallelic marker comprise one of the sequences selected from the group consisting of P1 to P28. Likewise, the term "marker" or "biallelic marker" requires that the sequence is of sufficient length to practically (although not necessarily unambiguously) identify the polymorphic allele, which usually implies a length of at least 4, 5, 6, 10, 15, 20, 25, or 40 nucleotides.

The invention also concerns *FLAP*-related biallelic markers. The term "*FLAP*-related biallelic marker" and "biallelic marker of the *FLAP* gene" are used interchangeably herein to relate to all biallelic markers in linkage disequilibrium with the *FLAP* gene. The term *FLAP*-related biallelic marker includes, but is not limited to, both the genic and non-genic biallelic markers described in Figure 1.

The term "non-genic" is used herein to describe *FLAP*-related biallelic markers, as well as polynucleotides and primers which occur outside the nucleotide positions shown in the human *FLAP* genomic sequence of SEQ ID No 1. The term "genic" is used herein to describe *FLAP*-related biallelic markers as well as polynucleotides and primers which do occur in the nucleotide positions shown in the human *FLAP* genomic sequence of SEQ ID No 1.

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to a disease involving the leukotriene pathway; or to refer to an individual's response to an agent acting on the leukotriene pathway; or to refer to symptoms of, or susceptibility to side-effects to an agent acting on the leukotriene pathway.

The term "disease involving the leukotriene pathway" refers to a condition linked to disturbances in expression, production or cellular response to leukotrienes. The diseases involving the leukotriene pathway include, but are not limited to, such as angina, endotoxic shock, psoriasis, atopic eczema, rheumatoid arthritis, inflammatory bowel disease, osteoarthritis, tendinitis, bursitis, ulcerative colitis, allergic bronchoasthma, allergic rhinitis, allergic conjunctivitis, glomerulonephritis, migraine headaches, and more particularly asthma.

The terms "response to an agent acting on the leukotriene pathway" refer to drug efficacy, including but not limited to ability to metabolize a compound, to the ability to convert a pro-drug to an active drug, and to the pharmacokinetics (absorption, distribution, elimination) and the pharmacodynamics (receptor-related) of a drug in an individual. In the context of the present invention, a "positive response" to a medicament can be defined as comprising a reduction of the symptoms related to the disease or condition to be treated. In the context of the present invention, a "negative response" to a medicament can be defined as comprising either a lack of positive response to the medicament which does not lead to a symptom reduction or to a side-effect observed following administration of the medicament.

The terms "side-effects to an agent acting on the leukotriene pathway" refer to adverse effects of therapy resulting from extensions of the principal pharmacological action of the drug or to idiosyncratic adverse reactions resulting from an interaction of the drug with unique host factors. The side-effects related to treatment with agents acting on the leukotriene pathway are preferably an acute exacerbation of an inflammatory disease such as asthma, infection and headache, and more preferably an increase in liver transaminase levels.

The terms "agents acting on the leukotriene pathway" preferably refer to a drug or a compound which modulates the activity or concentration of any enzyme or regulatory molecule involved in the leukotriene pathway in a cell or animal. Preferably these agents can be selected from the following group: FLAP inhibitors such as BAYx 1005, MK-886, and MK-0591; 5-Lipoxygenase inhibitors such as Zileuton, BAY-G576, RS-43,179, Wy-47,288, vitamin A, and BW A4C; Leukotriene LTD4 receptor antagonists such as zafirlukast, ICI 204,219, MK-571, MK-679, ONO-RS-411, SK&F 104,353, and Wy-48,252; Leukotriene B4 receptor antagonists; Leukotriene C4 synthase inhibitors; and, Leukotriene A4 hydrolase inhibitors. "Agents acting on the leukotriene pathway" further refers to non-steroidal antiinflammatory drugs (NSAIDs), leukotriene receptor

antagonists and leukotriene analogs. "Agents acting on the leukotriene pathway" also refers to compounds modulating the formation and action of leukotrienes.

Some of the compounds cited above are described in US patents 4,873,259; 4,970,215; 5,310,744; 5,225,421; and 5,081,138; or in EP 0 419 049, the disclosures of which are incorporated herein by reference.

The term "individual" as used herein refers to vertebrates, particularly members of the mammalian species and includes but is not limited to domestic animals, sports animals, laboratory animals, primates and humans. Preferably, an individual is a human.

Variants and fragments

Polynucleotides

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of a *FLAP* gene containing one or more biallelic markers according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Changes in the nucleotide of a variant may be silent, which means that they do not alter the amino acids encoded by the polynucleotide.

However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature FLAP protein.

A polynucleotide fragment is a polynucleotide having a sequence that entirely is the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of a *FLAP* gene, and variants thereof. The fragment can be a portion of an exon or of an intron of a *FLAP* gene. It can also be a portion of the regulatory sequences of the *FLAP* gene. Preferably, such fragments comprise the polymorphic base of at least one of the biallelic markers A1 to A28, the complement therefor, or a biallelic marker in linkage disequillibrium with one or more of the biallelic markers A1 to A28.

Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. However, several fragments may be comprised within a single larger polynucleotide.

As representative examples of polynucleotide fragments of the invention, there may be mentioned those which have from about 4, 6, 8, 15, 20, 25, 40, 10 to 20, 10 to 30, 30 to 55, 50 to 100, 75 to 100 or 100 to 200 nucleotides in length. Preferred are those fragments having about 47 nucleotides in length, such as those of P1 to P28, and containing at least one of the biallelic markers of a *FLAP* gene which are described herein. It will of course be understood that the polynucleotides P1 to P28 can be shorter or longer, although it is preferred that they at least contain the polymorphic base of the biallelic marker which can be located at one end of the fragment.

Polypeptides

The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein, including mutated FLAP proteins.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated FLAP is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated FLAP, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated FLAP or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

A polypeptide fragment is a polypeptide having a sequence that entirely is the same as part but not all of a given polypeptide sequence, preferably a polypeptide encoded by a *FLAP* gene and variants thereof. Preferred fragments include those of the active region of the FLAP protein that play a role in leukotriene biosynthesis and those regions possessing antigenic properties and which can be used to raise antibodies against the FLAP protein.

Such fragments may be "free-standing", i.e. not part of or fused to other polypeptides, or they may be comprised within a single larger polypeptide of which they form a part or region. However, several fragments may be comprised within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 5, 6, 7, 8, 9 or 10 to 15, 10 to 20, 15 to 40, or 30 to 55 amino acids long. Preferred are those fragments containing at least one amino acid mutation in the FLAP protein.

Stringent hybridization conditions

By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2 x SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1 X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989; and Ausubel et al., 1989, are incorporated herein in their entirety. These hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The suitable hybridization conditions may for example be adapted according to the teachings disclosed in the book of Hames and Higgins (1985) or in Sambrook et al.(1989).

Identity Between Nucleic Acids Or Polypeptides

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988; Altschul et al., 1990; Thompson et al., 1994; Higgins et al., 1996; Altschul et al., 1990; Altschul et al., 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul et al.,

1990; Altschul et al., 1993; Altschul et al., 1997). In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992; Henikoff and Henikoff, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990).

II. Genomic Sequences Of *FLAP*

Although the *FLAP* gene is of high relevance to pharmaceutical research, we still have scant knowledge concerning the extent and nature of sequence variation in this gene and its regulatory elements. The cDNA and part of the genomic sequence for human *FLAP* have been cloned and sequenced (Kennedy et al. 1991; Dixon et al, 1988). But, the complete genomic sequence of *FLAP*, including its regulatory elements, have not been described.

The present invention encompasses the genomic sequence of the *FLAP* gene of SEQ ID No 1 or a variant thereof or the complementary sequence thereto. This polynucleotide of nucleotide sequence of SEQ ID No 1, or a variant thereof or the complementary sequence thereto, may be purified, isolated, or recombinant. The *FLAP* genomic sequences comprise exons and introns. The nucleic acids derived from the *FLAP* intronic polynucleotides may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the *FLAP* gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the *FLAP* sequences.

The invention also encompasses a purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with a nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto or a fragment thereof. The nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 may be generally randomly distributed throughout the entire nucleic acid. Nevertheless, preferred nucleic acids are those wherein the nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 are predominantly located outside the coding sequences contained in the exons. These nucleic acids, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the *FLAP* gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the *FLAP* sequences.

The *FLAP* genomic nucleic acid comprises 5 exons. Exon 1 starts at the nucleotide in position 7709 and ends at the nucleotide in position 7852 of the nucleotide sequence of SEQ ID No 1; Exon 2 starts at the nucleotide in position 16236 and ends at the nucleotide in position 16335 of the nucleotide sequence of SEQ ID No 1; Exon 3 starts at the nucleotide in position 24227 and ends at the nucleotide in position 24297 of the nucleotide sequence of SEQ ID No 1; Exon 4 starts at the nucleotide in position 28133 and ends at the nucleotide in position 28214 of the nucleotide sequence of SEQ ID No 1; Exon 5 starts at the nucleotide in position 36128 and ends at the nucleotide in position 36605 of the nucleotide sequence of SEQ ID No 1. The invention also deals with purified, isolated, or recombinant nucleic acids comprising a combination of at least two exons of the *FLAP* gene, wherein the polynucleotides are arranged within the nucleic acid, from the 5'-end to the 3'-end of said nucleic acid, in the same order than in SEQ ID No 1.

The present invention also concerns a purified and/or isolated nucleic acid encoding a FLAP protein, preferably comprising at least one of the biallelic polymorphisms described herewith, and more preferably a *FLAP* gene comprising the trait-causing mutation determined using the below-noted method. In some embodiments, the *FLAP* gene comprises one or more of the sequences of P1 to P13, P15, and P17 to P28, or the complementary sequence thereto, or a fragment or a variant thereof. Preferred polynucleotides comprise at least one biallelic marker selected from the group consisting of A1 to A13, A15, A17 to A28, and the complements thereof. The present invention also provides polynucleotides which, may be used as primers and probes in order to amplify fragments carrying biallelic markers or in order to detect biallelic marker alleles.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069. Other preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35,

40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises a C at position 16348, of SEQ ID No 1. Further preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329. It should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section. Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1.

While this section is entitled "Genomic Sequences of *FLAP*," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of *FLAP* on either side or between two or more such genomic sequences.

Regulatory Regions Of The *FLAP* Gene

The genomic sequence of the *FLAP* gene contains regulatory sequences both in the non-coding 5'- flanking region and in the non-coding 3'- flanking region that border the *FLAP* transcribed region containing the 5 exons of this gene. 5'-regulatory sequences of the *FLAP* gene comprise the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 7708 of the nucleotide sequence of SEQ ID No 1, more preferably between positions 1 and 7007 of SEQ ID No 1. 3'-regulatory sequences of the *FLAP* gene comprise the polynucleotide sequences located between the nucleotide in position 36606 and the nucleotide in position 43069 of the nucleotide sequence of SEQ ID No 1.

Polynucleotides carrying the regulatory elements located both at the 5' end and at the 3' end of the *FLAP* coding region may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest, said polynucleotide being heterologous as regards to the *FLAP* regulatory region.

Thus, the present invention also concerns a purified, isolated, and recombinant nucleic acid comprising a polynucleotide which is selected from the group consisting of the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 7708 of the nucleotide sequence of SEQ ID No 1, more preferably between positions 1 and 7007 of SEQ ID No 1 and the polynucleotide sequences located between the nucleotide in position 36606 and the

nucleotide in position 43069 of SEQ ID No 1; or a sequence complementary thereto or a biologically active fragment thereof.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity, advantageously 99% nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 7708 of the nucleotide sequence of SEQ ID No 1, more preferably between positions 1 and 7007 of SEQ ID No 1 and the polynucleotide sequences located between the nucleotide in position 36606 and the nucleotide in position 43069 of SEQ ID No 1 or a variant thereof or a biologically active fragment thereof.

Another object of the invention consists of purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined therein, with a polynucleotide selected from the group consisting of the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 7007 of SEQ ID No 1 and the polynucleotide sequences located between the nucleotide in position 36606 and the nucleotide in position 43069 of SEQ ID No, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

Furthermore, the present invention also concerns a purified, isolated, and recombinant nucleic acid comprising a polynucleotide which is selected from the group consisting of:

- the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 7708 of the nucleotide sequence of SEQ ID No 1, more preferably between positions 1 and 7007 of SEQ ID No 1, and comprising a biallelic marker selected from the group consisting of A1 to A11 and A25 to A28, or a sequence complementary thereto or a biologically active fragment thereof; and
- the polynucleotide sequences located between the nucleotide in position 36606 and the nucleotide in position 43069 of SEQ ID No 1 and comprising a biallelic marker selected from the group consisting of A22 to A24 and the complements thereof, or a sequence complementary thereto or a biologically active fragment thereof.

By a "biologically active" fragment of SEQ ID No 1 according to the present invention is intended a polynucleotide comprising or alternatively consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are "operably linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

Preferred fragments of the 5'- or 3'-regulatory sequences have a length of about 1500 or 1000 nucleotides, preferably of about 500 nucleotides, more preferably about 400 nucleotides, even more preferably 300 nucleotides and most preferably about 200 nucleotides.

The regulatory polynucleotides of the invention may be prepared from the polynucleotide of SEQ ID No 1 by cleavage using suitable restriction enzymes, as described for example in the book of Sambrook et al.(1989). The regulatory polynucleotides may also be prepared by digestion of the polynucleotide of SEQ ID No 1 by an exonuclease enzyme, such as Bal31 (Wabiko et al., 1986). These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification.

The regulatory polynucleotides according to the invention may be advantageously part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism.

A preferred 5'-regulatory polynucleotide of the invention includes the 5'-untranslated region (5'-UTR) of the *FLAP* cDNA, or a biologically active fragment or variant thereof.

A preferred 3'-regulatory polynucleotide of the invention includes the 3'-untranslated region (3'-UTR) of the *FLAP* cDNA, or a biologically active fragment or variant thereof.

A further object of the invention consists of an isolated, purified or recombined polynucleotide comprising:

a) a nucleic acid comprising a regulatory nucleotide sequence selected from the group consisting of:

(i) a polynucleotide beginning at position 1 and ending at position 7708 of SEQ ID No 1, more preferably beginning at position 1 and ending at position 7007 of SEQ ID No 1, or a sequence complementary thereto;

(ii) a polynucleotide having at least 95% of nucleotide identity with the nucleotide sequence beginning at position 1 and ending at position 7708 of SEQ ID No 1, more preferably beginning at position 1 and ending at position 7007 of SEQ ID No 1, or a sequence complementary thereto;

(iii) a polynucleotide that hybridizes under stringent hybridization conditions with the nucleotide sequence beginning at position 1 and ending at position 7007 of SEQ ID No 1, or a sequence complementary thereto;

(iv) a biologically active fragment or variant of the polynucleotides in (i), (ii) and (iii);

b) a polynucleotide encoding a desired polypeptide or a nucleic acid of interest, operably linked to the nucleic acid defined in (a) above;

c) Optionally, a nucleic acid comprising a 3'- regulatory polynucleotide, preferably a 3'-regulatory polynucleotide of the *FLAP* gene.

In a specific embodiment of the nucleic acid defined above, said nucleic acid includes the 5'-untranslated region (5'-UTR) of the *FLAP* cDNA, or a biologically active fragment or variant thereof.

In a second specific embodiment of the nucleic acid defined above, said nucleic acid includes the 3'-untranslated region (3'-UTR) of the *FLAP* cDNA, or a biologically active fragment or variant thereof.

The regulatory sequences may comprise a biallelic marker selected from the group consisting of A1 to A11 and A22 to A28, and the complements thereof.

The polypeptide encoded by the nucleic acid described above may be of various nature or origin, encompassing proteins of prokaryotic or eukaryotic origin. Among the polypeptides expressed under the control of a *FLAP* regulatory region, there may be cited bacterial, fungal or viral antigens. Also encompassed are eukaryotic proteins such as intracellular proteins, for example "house keeping" proteins, membrane-bound proteins, for example receptors, and secreted proteins, for example cytokines. In a specific embodiment, the desired polypeptide may be the FLAP protein, especially the protein of the amino acid sequence of SEQ ID No 3.

The desired nucleic acids encoded by the above described polynucleotide, usually a RNA molecule, may be complementary to a desired coding polynucleotide, for example to the *FLAP* coding sequence, and thus useful as an antisense polynucleotide.

Such a polynucleotide may be included in a recombinant expression vector in order to express the desired polypeptide or the desired nucleic acid in host cell or in a host organism.

III. *FLAP* cDNA Sequences

The present invention provides a *FLAP* cDNA of SEQ ID No 2. The cDNA of SEQ ID No 2 also includes a 5'-UTR region and a 3'-UTR region. The 5'-UTR region stars at the nucleotide at position 1 and ends at the nucleotide in position 74 of SEQ ID No 2. The 3'-UTR region starts at the nucleotide at position 561 and ends at the nucleotide at position 875 of SEQ ID No 2. The polyadenylation site starts at the nucleotide at position 851 and ends at the nucleotide in position 856 of SEQ ID No 2.

Consequently, the invention concerns a purified, isolated, and recombinant nucleic acids comprising a nucleotide sequence of the 5'UTR and the 3'UTR of the *FLAP* cDNA, a sequence complementary thereto, or an allelic variant thereof.

Another object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID No 2, complementary sequences thereto or a variant or fragment thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant *FLAP* cDNAs consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 2. A particular preferred embodiment of the invention includes isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or a complementary sequence thereto, wherein said contiguous span comprises a T at position 197 (A13), an A at position 453 (A20), or a G at position 779 (A21) of SEQ ID No 2.

Most biallelic polymorphism represent silent nucleotide substitutions but biallelic marker A20 is associated with amino acid changes from valine to isoleucine in position 127 in the corresponding FLAP polypeptide.

The polynucleotide disclosed above that contains the coding sequence of the *FLAP* gene of the invention may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the *FLAP* gene of the invention or may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression.

While this section is entitled "*FLAP* cDNA Sequences," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of *FLAP* on either side or between two or more such genomic sequences.

Coding Regions

The *FLAP* open reading frame is contained in the corresponding mRNA of SEQ ID No 2. More precisely, the effective *FLAP* coding sequence (CDS) spans from the nucleotide in position 75 (first nucleotide of the ATG codon) to the nucleotide in position 560 (end nucleotide of the TGA codon)of the polynucleotide sequence of SEQ ID No 2. The present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, wherein said contiguous span includes a isoleucine residue at amino acid position 127 in SEQ ID No 3.

The above disclosed polynucleotide that contains the coding sequence of the *FLAP* gene may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the *FLAP* gene of the invention or in contrast the signals may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression and/or amplification.

IV. Polynucleotide Constructs

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment.

DNA Construct That Enables Directing Temporal And Spatial *FLAP* Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of the FLAP protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the *FLAP* genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to the *FLAP* nucleotide sequence of SEQ ID Nos 1 and 2, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in the 5'-regulatory sequence or in an exon of the *FLAP* genomic sequence or within the *FLAP* cDNA of SEQ ID No 2. In a preferred embodiment, the *FLAP* sequence comprises a biallelic marker of the present invention, preferably one of the biallelic markers A1 to A28.

The present invention embodies recombinant vectors comprising any one of the polynucleotides described in the present invention.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn110 for controlling the *FLAP* gene expression, such as described by Gossen et al.(1992, 1995) and Furth et al.(1994). Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the *FLAP* gene, said minimal promoter or said *FLAP* regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a FLAP polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor.

In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the *FLAP* genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the *FLAP* genomic sequence, and is located on the genome downstream the first *FLAP* nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker consists of the thymidine kinase (*tk*) gene (Thomas et al., 1986), the hygromycine beta gene (Te Riele et al., 1990), the *hprt* gene (Van der Lugt et al., 1991; Reid et al., 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada et al., 1993; Yagi et al.1990). Preferably, the positive selection marker is located within a *FLAP* exon sequence so as to interrupt the sequence encoding a FLAP protein. These replacement vectors are described, for example, by Thomas et al.(1986; 1987), Mansour et al.(1988) and Koller et al.(1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a *FLAP* regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986). The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique has been first described by Gu et al.(1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al.(1995), or by lipofection of the enzyme into the cells, such as described by Baubonis et al.(1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu et al.(1993) and Sauer et al.(1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu et al.(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the *FLAP* gene by homologous recombination is constructed in such a way that selectable markers are flanked

by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the *FLAP* sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are described by Zou et al.(1994).

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the *FLAP* genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the *FLAP* genomic sequence, and is located on the genome downstream of the first *FLAP* nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al.(1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result of the breeding of two transgenic animals, the first transgenic animal bearing the *FLAP*-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al.(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae et al.(1995).

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a *FLAP* genomic sequence or a *FLAP* cDNA sequence, and most preferably an altered copy of a *FLAP* genomic or cDNA sequence, within a predetermined

location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination). In a specific embodiment, the DNA constructs described above may be used to introduce a *FLAP* genomic sequence or a *FLAP* cDNA sequence comprising at least one biallelic marker of the present invention, preferably at least one biallelic marker selected from the group consisting of A1 to A28 and the complements thereof, more preferably at least one biallelic marker selected from the group consisting of A1 to A13, A15, and A17 to A28 and the complements thereof.

Nuclear Antisense DNA Constructs

Other compositions containing a vector of the invention comprising an oligonucleotide fragment of the nucleic sequence SEQ ID No 2 comprising a biallelic marker of the invention, preferably a fragment including the start codon of the *FLAP* gene, as an antisense tool that inhibits the expression of the corresponding *FLAP* gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995) or those described in PCT Application No WO 95/24223.

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the *FLAP* mRNA. In one embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of *FLAP* that contains either the translation initiation codon ATG or a splicing site. Further preferred antisense polynucleotides according to the invention are complementary of the splicing site of the *FLAP* mRNA.

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu et al.(1994). In a preferred embodiment, these *FLAP* antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner et al.(1991).

V. Biallelic Markers Of The *FLAP* Gene

The invention also concerns *FLAP*-related biallelic markers, preferably a biallelic marker associated with a disease involving the leukotriene pathway, most preferably asthma. The term *FLAP*-related biallelic marker includes the biallelic markers designated A1 to A28. The invention also concerns sets of these biallelic markers.

28 biallelic markers have been identified in the genomic sequence of *FLAP*. These biallelic markers are disclosed in Table 2 of Example 3. Their location on the *FLAP* genomic sequence and cDNA is indicated in Table 2 and also as a single base polymorphism in the features of SEQ ID No 1. The Table 2 also disclosed the position on the SEQ ID No 1 of polynucleotides of 47 nucleotides in length, designated P1 to P28, which comprise a biallelic marker of the *FLAP* gene and define said biallelic marker. The pairs of primers allowing the amplification of a nucleic acid containing the polymorphic base of one *FLAP* biallelic marker are listed in Table 1 of Example 2. Three biallelic markers, namely A13, A20 and A21, are located in exonic regions. Two of them do not modify the amino acid sequence of the FLAP protein. However, the biallelic marker A20 changes a valine into a isoleucine in the FLAP protein.

The invention also relates to a purified and/or isolated nucleotide sequence comprising a polymorphic base of a biallelic marker located in the sequence of the *FLAP* gene, preferably of a biallelic marker selected from the group consisting of A1 to A28, preferably from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof; optionally, said biallelic marker is selected from the group consisting of A1 to A10 and A22 to A28; optionally, said biallelic marker is selected from the group consisting of A11 to A13, A15, A17 to A21; optionally, said biallelic marker is either A14 or A16. The sequence has between 8 and 1000 nucleotides in length, and preferably comprises at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a variant thereof or a complementary sequence thereto. These nucleotide sequences comprise the polymorphic base of either allele 1 or allele 2 of the considered biallelic marker. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of said polynucleotide or at the center of said polynucleotide. Optionally, the 3' end of said contiguous span may be present at the 3' end of said polynucleotide. Optionally, biallelic marker may be present at the 3' end of said polynucleotide. Optionally, the 3' end of said polynucleotide may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a biallelic marker of the *FLAP* gene in said sequence. Optionally, the 3' end of said polynucleotide may be located 1 nucleotide upstream of a biallelic marker of the *FLAP* gene in said sequence. Optionally, said polynucleotide may further comprise a label. Optionally, said polynucleotide can be attached to solid support. In a further embodiment, the polynucleotides defined above can be used alone or in any combination.

The invention further concerns a nucleic acid encoding the *FLAP* protein, wherein said nucleic acid comprises a polymorphic base of a biallelic marker selected from the group consisting of A1 to A28 and the complements thereof, preferably from the group consisting of A1 to A13, A15, and A17 to A28 and the complements thereof.

The invention also relates to a nucleotide sequence, preferably a purified and/or isolated nucleotide sequence comprising a sequence defining a biallelic marker of the *FLAP* gene, a fragment

or variant thereof or a sequence complementary thereto, said fragment comprising the polymorphic base. Preferably, the sequences defining a biallelic marker include the polymorphic base of one of the polynucleotides P1 to P13, P15 and P17 to P28 or the complements thereof. In some embodiments, the sequences defining a biallelic marker comprise a nucleotide sequence selected from the group consisting of P1 to P13, P15 and P17 to P28, and the complementary sequence thereto or a fragment thereof, said fragment comprising the polymorphic base.

The invention also concerns a set of the purified and/or isolated nucleotide sequences defined above. More particularly, the set of purified and/or isolated nucleotide sequences comprises a group of sequences defining a combination of biallelic markers of the *FLAP* gene. Preferably, the combination of alleles of biallelic markers is associated with asthma.

In a preferred embodiment, the invention relates to a set of purified and/or isolated nucleotide sequences, each sequence comprising a sequence defining a biallelic marker of the *FLAP* gene, wherein the set is characterized in that between about 30 and 100%, preferably between about 40 and 60 %, more preferably between 50 and 60%, of the sequences defining a biallelic marker are selected from the group consisting of P1 to P28, preferably of P1 to P13, P15 and P17 to P28, or a fragment or variant thereof or the complementary sequence thereto, said fragment comprising the polymorphic base.

More particularly, the invention concerns a set of purified and/or isolated nucleotide sequences, each sequence comprising a sequence defining a different biallelic marker of the *FLAP* gene, said biallelic marker being either included in a nucleotide sequence selected from the group consisting of P1 to P28 and the complementary sequence thereto, preferably of P1 to P13, P15 and P17 to P28 and the complementary sequence thereto, or a biallelic marker, preferably one located in the sequence of the *FLAP* gene, biallelic markers A1 to A 28, or markers in linkage disequilibrium with one of the markers of the set defined herewith.

The invention also relates to a set of at least two, preferably four, five, six, seven, eight or more nucleotide sequences selected from the group consisting of P1 to P28, preferably of P1 to P13, P15 and P17 to P28, and the complementary sequence thereto, or a fragment or variant thereof, said fragment comprising the polymorphic base. Preferably, this set comprises at least one nucleotide sequence defining a biallelic marker for each linkage disequilibrium region of the *FLAP* gene.

The invention further concerns a nucleotide sequence selected from the group consisting of P1 to P13, P15 and P17 to P28, or a complementary sequence thereto or a fragment or a variant thereof, said fragment comprising the polymorphic base.

In a further embodiment, the sequences comprising a polymorphic base of one of the biallelic markers listed in Table 2 are selected from the group consisting of the nucleotide sequences that have a contiguous span of, that consist of, that are comprised in, or that comprises a polynucleotide selected from the group consisting of the nucleic acids of the sequences set forth as Nos. 10-517, 10-518, 10-253, 10-499, 10-500, 10-522, 10-503, 10-504, 10-204, 10-32, 10-33, 10-34,

10-35, 10-36, 10-498, 12-628, and 12-629 (listed in Table 1) or a variant thereof or a complementary sequence thereto.

VI. Oligonucleotide Probes and Primers

Polynucleotides derived from the *FLAP* gene are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID No 1 or 2, or a fragment or a variant thereof in a test sample.

Particularly preferred probes and primers comprise a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069. Other preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises a C at position 16348, of SEQ ID No 1. Further preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 26, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises of the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329. Additional preferred probes and primers comprise a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1.

Thus, the invention also relates to nucleic acid probes or primers characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a nucleic acid selected from the group consisting of the nucleotide sequences 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069 of SEQ ID No 1 or a variant thereof or a sequence complementary thereto.

Particularly preferred probes and primers comprise a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or a complementary sequence thereto, wherein said contiguous span comprises a T at position 197 (A13), an A at position 453 (A20), or a G at position 779 (A21) of SEQ ID No 2.

The present invention also concerns oligonucleotides and groups of oligonucleotides for the detection of alleles associated with a modified leukotriene metabolism, preferably alleles associated with a *FLAP* gene polymorphism, and more preferably alleles of a *FLAP* gene associated with a

disease involving the leukotriene pathway, for example asthma. These oligonucleotides are characterized in that they can hybridize with a *FLAP* gene, preferably with a polymorphic *FLAP* gene and more preferably with a region of a *FLAP* gene comprising the polymorphic site of which specific alleles are associated with a disease involving the leukotriene pathway such as asthma. The oligonucleotides are useful either as primers for use in various processes such as DNA amplification and microsequencing or as probes for DNA recognition in hybridization analyses. In some embodiments, the oligonucleotides contain the polymorphic base of a sequence selected from the group consisting of P1 to P28 and the complementary sequence thereto, more preferably from the group consisting of P1 to P13, P15, P17 to P28 and the complementary sequence thereto. In other embodiments, the oligonucleotides have a 3' terminus immediately adjacent to a polymorphic base in the *FLAP* gene, such as a polymorphic base in one of P1 to P28 and the complementary sequence thereto, optionally of P1 to P13, P15, and P17 to P28 and the complementary sequence thereto. In other embodiments, the oligonucleotide is capable of discriminating between different alleles of a biallelic marker in the *FLAP* gene, said biallelic marker being selected from the group consisting of A1 to A28 and the complements thereof, optionally of A1 to A13, A15, and A17 to A28 and the complements thereof. For example, the oligonucleotide may be capable of specifically hybridizing to one allele of a biallelic marker, including one of the biallelic markers A1 to A28 and the complements thereof, optionally of A1 to A13, A15, and A17 to A28 and the complements thereof. In another embodiment, the oligonucleotides comprise one of the sequences of B1 to B17, C1 to C17, D1 to D28, E1 to E28, and P1 to P28, and the complementary sequence thereto. Optionally, the oligonucleotides comprise one of the sequences of B1 to B17, C1 to C17, D1 to D13, D15, D17 to D28, E1 to E13, E15, E17 to E28, P1 to P13, P15, and P17 to P28, and the complementary sequence thereto.

In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 or 2 and the complement thereof, wherein said span includes a *FLAP*-related biallelic marker in said sequence; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A13, A15, A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A14 and A16, and the complements thereof, or optionally the biallelic

markers in linkage disequilibrium therewith; optionally, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the polynucleotide and preferably within 4 nucleotides of the center of said polynucleotide; optionally, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide; optionally, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

In another embodiment the invention encompasses isolated, purified and recombinant polynucleotides consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 or 2 or the complement thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide. In one embodiment, the 3' end of said polynucleotide is located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a biallelic marker of *FLAP* in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers. In a particular embodiment, the 3' end of said polynucleotide is located within 20 nucleotides upstream of a *FLAP*-related biallelic marker in said sequence; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A13, A15, A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said *FLAP*-related biallelic marker is either A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said *FLAP*-related biallelic marker in said sequence; and optionally, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D28 and E1 to E28; and optionally, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D13, D15, D17 to D28, E1 to E13, E15, and E17 to E28. In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides consisting of, or consisting essentially of a sequence selected from the following sequences: B1 to B17, and C1 to C17. To these primers can be added, at either end thereof, a further polynucleotide useful for sequencing. Preferably, primers PU contain

the additional PU 5' sequence of SEQ ID No 14 and primers RP contain the RP 5' sequence of SEQ ID No 15.

In an additional embodiment, the invention encompasses polynucleotides for use in hybridization assay, sequencing assays, microsequencing assays and enzyme-based mismatch detection assays for determining the identity of the nucleotide at a *FLAP*-related biallelic marker in SEQ ID No 1 or 2, or the complement thereof, as well as polynucleotides for use in amplifying segments of nucleotides comprising a *FLAP*-related biallelic marker in SEQ ID No 1 or 2, or the complement thereof; optionally, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally from the group consisting of A14 and A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification. A preferred polynucleotide may be used in a hybridization assay for determining the identity of the nucleotide at a biallelic marker of the *FLAP* gene. Another preferred polynucleotide may be used in a sequencing or microsequencing assay for determining the identity of the nucleotide at a biallelic marker of the *FLAP* gene. A third preferred polynucleotide may be used in an enzyme-based mismatch detection assay for determining the identity of the nucleotide at a biallelic marker of the *FLAP* gene. A fourth preferred polynucleotide may be used in amplifying a segment of polynucleotides comprising a biallelic marker of the *FLAP* gene; Optionally, said amplifying may be performed by a PCR or LCR. Optionally, said polynucleotide may be attached to a solid support, array, or addressable array; Optionally, said polynucleotide may be labeled.

Primers and probes according to the invention are therefore synthesized to be "substantially" complementary to a strand of the *FLAP* gene to be amplified. The primer sequence does not need to reflect the exact sequence of the DNA template. Minor mismatches can be accommodated by reducing the stringency of the hybridization conditions. Among the various methods available to design useful primers, the OSP computer software can be used by the skilled person (see Hillier & Green, 1991).

The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature

because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the primers and probes of the invention usually ranges between 10 and 75%, preferably between 35 and 60%, and more preferably between 40 and 55%.

Preferably, the length of the primer and probe can range from 10 to 100 nucleotides, preferably from 10 to 50, 10 to 30 or more preferably 10 to 25 nucleotides. Shorter primers and probes tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer primers and probes are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art.

The probes of the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the *FLAP* gene or mRNA using other techniques. Generally, the probes are complementary to the *FLAP* gene coding sequences, Although probes to introns and regulatory sequences are also contemplated.

Primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang et al. (1979), the phosphodiester method of Brown et al. (1979), the diethylphosphoramidite method of Beaucage et al. (1981) and the solid support method described in EP 0 707 592. The disclosures of all these documents are incorporated herein by reference.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702; morpholino analogs which are described in U.S. Patents Numbered 5,185,444, 5,034,506, and 5,142,047; and the like. The disclosures of each of these patents is incorporated herein by reference. Depending upon the type of label carried by the probe, the probe is employed to capture or detect the amplicon generated by the amplification reaction. The probe is not involved in amplification of the target sequence and therefore may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified. U.S. Patent Application Serial No 07/049,061 filed April 19, 1993 describes modifications which can be used render a probe non-extendable.

The probes are preferably directly labeled such as with isotopes, reporter molecules or fluorescent labels or indirectly labeled such as with biotin to which a streptavidin complex may later bind. Probe labeling techniques are well-known to the skilled technician. By assaying the presence of the probe, one can detect the presence or absence of the targeted DNA sequence in a given sample. The same labels can be used with primers. For example, useful labels include radioactive substances (^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent dyes (5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin). Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in the French patent No FR-7810975 or by Urdea et al (1988) or Sanchez-Pescador et al (1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European patent No EP 0 225 807 (Chiron).

Any of the primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The "solid phase" is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples.

Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A "solid phase", as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid phase can be chosen for its intrinsic ability to attract and immobilize the capture reagent.

Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent.

As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid phase and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid phase material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art.

The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Consequently, the invention also deals with a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1 and 2, a fragment or a variant thereof or a complementary sequence thereto in a sample, said method comprising the following steps of:

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected form the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2, a fragment or a variant thereof or a complementary sequence thereto and the sample to be assayed.

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1 and 2, a fragment or a variant thereof or a complementary sequence thereto in a sample, said kit comprising:

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected form the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2, a fragment or a variant thereof or a complementary sequence thereto;

b) optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of the detection method and kit, the nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule. In a second preferred embodiment of the detection method and kit, the nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate. In a third preferred embodiment of the detection method and kit, the nucleic acid probe or the plurality of nucleic acid probes comprise either a sequence which is selected from the group consisting of the nucleotide sequences: B1 to B17, C1 to C17, D1 to D28, E1 to E28, P1 to P28 or a biallelic marker selected from the group consisting of A1 to A28 or the complements thereto or the biallelic markers in linkage disequilibrium therewith.

Oligonucleotide Arrays

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in the *FLAP* gene and may also be used for detecting mutations in the coding or in the non-coding sequences of the *FLAP* gene.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an

ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in the *FLAP* gene. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the *FLAP* gene that have been identified according, for example to the technique used by Huang et al.(1996) or Samson et al.(1996).

Another technique that is used to detect mutations in the *FLAP* gene is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the *FLAP* genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the *FLAP* gene. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement

will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. This technique was described by Chee et al. in 1996, which is herein incorporated by reference.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

A further object of the invention consists of an array of nucleic acid sequences comprising either at least one of the sequences selected from the group consisting of P1 to P28, B1 to B17, C1 to C17, D1 to D28 and E1 to E28 or the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 30, or 40 consecutive nucleotides thereof, or at least one sequence comprising a biallelic marker selected from the group consisting of A1 to A28, and the complements thereto, or optionally the biallelic markers in linkage disequilibrium therewith.

The invention also pertains to an array of nucleic acid sequences comprising either at least two of the sequences selected from the group consisting of P1 to P28, B1 to B17, C1 to C17, D1 to D28 and E1 to E28 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides thereof, or at least two sequences comprising a biallelic marker selected from the group consisting of A1 to A28, and the complements thereto, or optionally the biallelic markers in linkage disequilibrium therewith.

VII. Identification Of Biallelic Markers

There are two preferred methods through which the biallelic markers of the present invention can be generated. In a first method, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms.

One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained therewith usually shows a sufficient degree of informativeness for conducting association studies.

In a second method for generating biallelic markers, the DNA samples are not pooled and are therefore amplified and sequenced individually. The resulting nucleotide sequences obtained are then also analyzed to identify significant polymorphisms.

It will readily be appreciated that when this second method is used, a substantially higher number of DNA amplification reactions and sequencing reactions must be carried out. Moreover, a

biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. It will further be appreciated that including such less informative biallelic markers in association studies to identify potential genetic associations with a trait may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes.

The following is a description of the various parameters of a preferred method used by the inventors to generate the markers of the present invention.

1. DNA extraction

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background.

The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples which can be tested by the methods of the present invention described herein and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the context of the present invention is from peripheral venous blood of each donor.

The techniques of DNA extraction are well-known to the skilled technician. Details of a preferred embodiment are provided in Example 1.

Once genomic DNA from every individual in the given population has been extracted, it is preferred that a fraction of each DNA sample is separated, after which a pool of DNA is constituted by assembling equivalent amounts of the separated fractions into a single one. However, the person skilled in the art can choose to amplify the pooled or unpooled sequences

2. DNA amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step.

DNA amplification techniques are well-known to those skilled in the art. Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the disclosures of which are incorporated herein by reference, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli J.C., et al.(1990) and in Compton J.(1991), Q-beta amplification as described in European Patent Application No 4544610, strand displacement amplification as described in Walker et al.(1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461, the disclosures of which are incorporated herein by reference.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes, which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall et al.(1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR

technology, see White (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188. Each of these publications is incorporated herein by reference.

The PCR technology is the preferred amplification technique used to identify new biallelic markers. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 2.

One of the aspects of the present invention is a method for the amplification of the human *FLAP* gene, particularly of the genomic sequence of SEQ ID No 1 or of the cDNA sequence of SEQ ID No 2, or a fragment or a variant thereof in a test sample, preferably using the PCR technology. This method comprises the steps of contacting a test sample suspected of containing the target *FLAP* encoding sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers, and eventually in some instances a detection probe that can hybridize with an internal region of amplicon sequences to confirm that the desired amplification reaction has taken place.

Thus, the present invention also relates to a method for the amplification of a human *FLAP* gene sequence, particularly of a portion of the genomic sequences of SEQ ID No 1 or of the cDNA sequence of SEQ ID No 2, or a variant thereof in a test sample, said method comprising the steps of:

a) contacting a test sample suspected of containing the targeted *FLAP* gene sequence comprised in a nucleotide sequence selected from a group consisting of SEQ ID Nos 1 and 2, or fragments or variants thereof with amplification reaction reagents comprising a pair of amplification primers as described above and located on either side of the polynucleotide region to be amplified, and

b) optionally, detecting the amplification products.

The invention also concerns a kit for the amplification of a human *FLAP* gene sequence, particularly of a portion of the genomic sequence of SEQ ID No 1 or of the cDNA sequence of SEQ ID No 2, or a variant thereof in a test sample, wherein said kit comprises:

a) a pair of oligonucleotide primers located on either side of the *FLAP* region to be amplified;

b) optionally, the reagents necessary for performing the amplification reaction.

In one embodiment of the above amplification method and kit, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In another embodiment of the above amplification method and kit, primers comprise a sequence which is selected from the group consisting of the nucleotide sequences of B1 to B17, C1 to C17, D1 to D28 and E1 to E28. In a preferred embodiment of the above amplification method and kit, the amplification product comprises a polymorphic base of a biallelic marker of the present invention, more particularly a polymorphic base of a biallelic marker selected from the group of A1 to A28, optionally from the group consisting of A1 to A13, A15 and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. The primers are more particularly characterized in that they have sufficient complementarity with any sequence of a strand of the genomic sequence close to the region to be amplified, for example with a non-coding sequence adjacent to exons to amplify.

In a first embodiment of the present invention, biallelic markers are identified using genomic sequence information generated by the inventors. Sequenced genomic DNA fragments are used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

Preferred primers, useful for the amplification of genomic sequences encoding the candidate genes, focus on promoters, exons and splice sites of the genes. A biallelic marker presents a higher probability to be an eventual causal mutation if it is located in these functional regions of the gene. Preferred amplification primers of the invention include the nucleotide sequences B1 to B17 and the nucleotide sequences C1 to C17 disclosed in Example 2.

3. Sequencing of amplified genomic DNA and identification of polymorphisms

The amplification products generated as described above with the primers of the invention are then sequenced using methods known and available to the skilled technician. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol.

Following gel image analysis and DNA sequence extraction, sequence data are automatically processed with software to assess sequence quality.

The sequence data obtained as described above are subjected to quality control and validation steps based on the shape of the peak, the inter-peak resolution, the number of unreliable peaks in a particular stretch of sequence and the noise level. Sequence data that is considered unreliable is discarded.

After this first sequence quality analysis, polymorphisms are detected among individual or pooled amplified fragment sequences. The polymorphism search is based on the presence of

superimposed peaks in the electrophoresis pattern. These peaks, which present two distinct colors, correspond to two different nucleotides at the same position on the sequence. In order for peaks to be considered significant, peak height has to satisfy conditions of ratio between the peaks and conditions of ratio between a given peak and the surrounding peaks of the same color.

However, since the presence of two peaks can be an artifact due to background noise, two controls are utilized to exclude these artifacts:

- the two DNA strands are sequenced and a comparison between the peaks is carried out. The polymorphism has to be detected on both strands for validation.

- all the sequencing electrophoresis patterns of the same amplification product provided from distinct pools and/or individuals are compared. The homogeneity and the ratio of homozygous and heterozygous peak height are controlled through these distinct DNAs.

The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is about 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90 % of the biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele, preferably at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than 0.7 for the major allele, thus a heterozygosity rate higher than 0.18, preferably higher than 0.32, more preferably higher than 0.42.

In another embodiment, biallelic markers are detected by sequencing individual DNA samples, the frequency of the minor allele of such a biallelic marker may be less than 0.1.

4. Validation of the biallelic markers of the present invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence.

5. Evaluation of the frequency of the biallelic markers of the present invention

The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The higher the frequency of the less common allele the greater the usefulness of the biallelic marker is in association studies. The determination of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. This determination of frequency by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. For an indication of the frequency for the less common allele of a particular biallelic marker of the invention see Table 2. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker."

The invention also relates to methods of estimating the frequency of an allele of a *FLAP*-related biallelic marker in a population comprising: a) genotyping individuals from said population for said biallelic marker according to the method of the present invention; b) determining the proportional representation of said biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A28, and the complements thereof; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, determining the proportional representation of a nucleotide at a *FLAP*-related biallelic marker may be accomplished by determining the identity of the nucleotides for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said nucleotide at said *FLAP*-related biallelic marker for the population; Optionally, determining the proportional representation may be accomplished by performing a genotyping method of the invention on a pooled biological sample

derived from a representative number of individuals, or each individual, in said population, and calculating the proportional amount of said nucleotide compared with the total.

VIII. Methods For Genotyping An Individual For Biallelic Markers

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at a *FLAP* biallelic marker site by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed on nucleic acid samples derived from a single individual or pooled DNA samples.

The identification of biallelic markers described previously allows the design of appropriate oligonucleotides, which can be used as probes and primers, to amplify a *FLAP* gene containing the polymorphic site of interest and for the detection of such polymorphisms.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

The invention also pertains to a method of genotyping comprising determining the identity of a nucleotide at a biallelic marker of the *FLAP* gene in a biological sample. Optionally, the biological sample is derived from a single subject; Optionally, the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome. Optionally, said method is performed in vitro; Optionally, the biological sample is derived from multiple subjects. Optionally, the method of genotyping described above further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; Optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said portion in a host cell. The determining step of the above genotyping method may be performed either by a hybridization assay, a sequencing assay, an enzyme-based mismatch detection assay and by a microsequencing assay. Thus, the invention also encompasses methods of genotyping a biological sample comprising determining the identity of a nucleotide at a *FLAP*-related biallelic marker. In addition, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. Optionally, said biallelic marker is selected from

the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic marker is selected from the group consisting of A1 to A13, A15 ,and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic marker is selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic marker is selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic marker is either A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith.

Source of DNA for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

Amplification of DNA fragments comprising biallelic markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention. Amplification of DNA may be achieved by any method known in the art. Amplification techniques are described above in the section entitled, "Identification of biallelic markers" VII. (2).

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described below.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. In some embodiments, the primer pair is adapted for amplifying a sequence containing the polymorphic base of one of the sequences of P1 to P28, optionally P1 to P13, P15, P17 to P28, and the complementary sequence thereto. Preferred amplification primers are listed in Example 2. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention, amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. In a preferred embodiment of the invention, the pairs of primers for amplification and sequencing are sufficiently complementary with a region of a *FLAP* gene located at less than 500 bp, preferably at less than 100 bp, and more preferably at less than 50 bp of a polymorphic site corresponding to one of the markers of the present invention. Amplification primers may be labeled or immobilized on a solid support as described in "Oligonucleotide probes and primers".

Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al.(1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield et al.(1991), White et al.(1992), Grompe et al.(1989 and 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by a sequencing assay, an enzyme-based mismatch detection assay, or a hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing Assays

The amplification products generated above with the primers of the invention can be sequenced using methods known and available to the skilled technician. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. A sequence analysis can allow the identification of the base present at the polymorphic site.

2) Microsequencing Assays

Polymorphism analyses on pools or selected individuals of a given population can be carried out by conducting microsequencing reactions on candidate regions contained in amplified fragments obtained by PCR performed on DNA or RNA samples taken from these individuals.

To do so, DNA samples are subjected to PCR amplification of the candidate regions under conditions similar to those described above. These genomic amplification products are then subjected to automated microsequencing reactions using ddNTPs (specific fluorescence for each ddNTP) and appropriate oligonucleotide microsequencing primers which can hybridize just upstream of the polymorphic base of interest. Once specifically extended at the 3' end by a DNA polymerase using a complementary fluorescent dideoxynucleotide analog (thermal cycling), the primer is precipitated to remove the unincorporated fluorescent ddNTPs. The reaction products in which fluorescent ddNTPs have been incorporated are then analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated base, thereby identifying the polymorphic marker presnet in the sample.

An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in example 4. It is to be understood that certain parameters of this procedure such as the electrophoresis method or the labeling of ddNTPs could be modified by the skilled person without substantially modifying its result.

The extended primer may also be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997).

As a further alternative to the process described above, several solid phase microsequencing reactions have been developed. The basic microsequencing protocol is the same as described previously, except that either the oligonucleotide microsequencing primers or the PCR-amplified products of the DNA fragment of interest are immobilized. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles.

In such solid phase microsequencing reactions, incorporated ddNTPs can either be radiolabeled (see Syvänen, 1994, incorporated herein by reference) or linked to fluorescein (see Livak & Hainer, 1994, incorporated herein by reference). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can

be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate).

Other possible of reporter-detection couples include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (see Harju et al., 1993, incorporated herein by reference); and biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (see WO 92/15712, incorporated herein by reference).

A diagnosis kit based on fluorescein-linked ddNTP with antifluorescein antibody conjugated with alkaline phosphatase is commercialized under the name PRONTO by GamidaGen Ltd.

As yet another alternative microsequencing procedure, Nyren et al. (1993) presented a concept of solid-phase DNA sequencing that relies on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA). The PCR-amplified products are biotinylated and immobilized on beads. The microsequencing primer is annealed and four aliquots of this mixture are separately incubated with DNA polymerase and one of the four different ddNTPs. After the reaction, the resulting fragments are washed and used as substrates in a primer extension reaction with all four dNTPs present. The progress of the DNA-directed polymerization reactions are monitored with the ELIDA. Incorporation of a ddNTP in the first reaction prevents the formation of pyrophosphate during the subsequent dNTP reaction. In contrast, no ddNTP incorporation in the first reaction gives extensive pyrophosphate release during the dNTP reaction and this leads to generation of light throughout the ELIDA reactions. From the ELIDA results, the first base after the primer is easily deduced.

Pastinen et al.(1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferably, the biallelic markers are selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally, the biallelic markers are selected from the group consisting of A1 to A13, A15, A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Preferred microsequencing primers include the nucleotide sequences: D1 to D28 and E1 to E28. Optionally, microsequencing primers include the nucleotide sequences: D1 to D13, D15, D17 to D28, E1 to E13, E15, and E17 to E28. More preferred microsequencing primers are selected from the group consisting of the nucleotide sequences: E11, D12, D13, D14, D15, D16, E18, D19, and E20. It will be appreciated that the microsequencing primers listed in Example 4 are merely exemplary and that, any primer having a 3' end immediately adjacent to the polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for

any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Example 4, or fragments comprising at least 8, 12, 15, 20, 25, 30, 40, or 50 consecutive nucleotides thereof and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at a biallelic marker site.

3) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by allele-specific amplification assays. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in "Amplification Of DNA Fragments Comprising Biallelic Markers".

Allele Specific Amplification Primers

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. For allele specific amplification, at least one member of the pair of primers is sufficiently complementary with a region of a *FLAP* gene comprising the polymorphic base of a biallelic marker of the present invention to hybridize therewith. Such primers are able to discriminate between the two alleles of a biallelic marker.

This can be accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Determining the precise location of the mismatch and the corresponding assay conditions are well with the ordinary skill in the art.

Ligation/Amplification Based Methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and may be advantageously combined with PCR as described by Nickerson et al.(1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other amplification methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in " Identification Of Biallelic Markers " (2). LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

4) Hybridization Assay Methods

The invention also relates to a group of probes characterized in that they preferably comprise between 10 and 50 nucleotides, and in that they are sufficiently complementary to a polymorphic sequence defined by a biallelic marker located in the genomic sequence of a *FLAP* gene to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation.

The length of these probes can range from 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 40 to 50 nucleotides. A particularly preferred probe is 25 nucleotides in length. An other preferred probe is 47 nucleotides in length. It includes a central nucleotide complementary to a polymorphic site of the *FLAP* gene, preferably a polymorphic site corresponding to one of the biallelic markers of the present invention, and a 23 nucleotide sequence spanning on each side of the central nucleotide and substantially complementary to the nucleotide sequences of the *FLAP* gene spanning on each side of the polymorphic site. Optionally, the biallelic markers of the present invention comprise the polymorphic bases in the sequences of P1 to P28 and

the complementary sequences thereto. Optionally, the biallelic markers of the present invention comprise the polymorphic bases in the sequences of P1 to P13, P15, and P17 to P28, and the complementary sequences thereto.

Polymorphisms can be analyzed and the frequency of corresponding alleles quantified through hybridization reactions on amplified *FLAP* encoding sequences. The amplification reaction can be carried out as described previously. The hybridization probes which can be conveniently used in such reactions preferably include the probes defined above as being sufficiently complementary to a polymorphic site defined by one of the biallelic markers located in the genomic sequence of a *FLAP* gene to hybridize thereto and sufficiently specific to be able to discriminate between the targeted allele and an allele differing by only one base.

The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., 1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., 1995). In an alternative homogeneous hybridization based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., 1998).

5) Hybridization To Addressable Arrays Of Oligonucleotides

Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA

probes arranged in a grid-like pattern and miniaturized to the size of a dime. These DNA chips are detailed in "oligonucleotides primers and probes", section "Oligonucleotide array".

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (see Hacia et al., 1996; Shoemaker et al., 1996; Kozal et al., 1996, incorporated herein by reference).

At least, three companies propose chips able to detect biallelic polymorphisms: Affymetrix (GeneChip), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

One of the limitations encountered when using DNA chip technology is that hybridization of nucleic acids with the probes attached to the chip in arrays is not simply a solution-phase reaction. A possible improvement consists in using polyacrylamide gel pads isolated from one another by hydrophobic regions in which the DNA probes are covalently linked to an acrylamide matrix.

For the detection of polymorphisms, probes which contain at least a portion of one of the biallelic markers of the present invention, such as the biallelic markers of P1 to P28, optionally P1 to P13, P15, and P17 to P28, and the complementary sequences thereto, are synthesized either *in situ* or by conventional synthesis and immobilized on an appropriate chip using methods known to the skilled technician. The solid surface of the chip is often made of silicon or glass but it can be a polymeric membrane. Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments thereof at least 15 nucleotides in length, preferably at least 20 nucleotides in length, and more preferably at least 25 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of P1 to P28, D1 to D28, and E1 to E28, or the sequences complementary thereto, or a fragment thereof at least 15 consecutive nucleotides. Optionally, the chip may comprise an array including at least one of the sequences selected from the group consisting of P1 to P13, P15, P17 to P28, D1 to D13, D15, D17 to D28, E1 to E13, E15, and E17 to E28, or the sequences complementary thereto, or a fragment thereof at least 15 consecutive nucleotides. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more sequences selected from the group consisting of P1 to P28, D1 to D28, and E1 to E28, or the sequences complementary thereto, or a fragment thereof at least 15 consecutive nucleotides. Optionally, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more sequences selected from the group consisting of P1 to P13, P15, P17 to P28, D1 to D13, D15, D17 to D28, E1 to E13, E15, and E17 to E28, or the sequences complementary thereto, or a fragment thereof at least 15 consecutive nucleotides.

The nucleic acid sample which includes the candidate region to be analyzed is isolated, amplified and labeled with a reporter group. This reporter group can be a fluorescent group such as phycoerythrin. The labeled nucleic acid is then incubated with the probes immobilized on the chip using a fluidics station. For example, Manz et al. (1993) describe the fabrication of fluidics devices and particularly microcapillary devices, in silicon and glass substrates.

After the reaction is completed, the chip is inserted into a scanner and patterns of hybridization are detected. The hybridization data is collected, as a signal emitted from the reporter groups already incorporated into the nucleic acid, which is now bound to the probes attached to the chip. Probes that perfectly match a sequence of the nucleic acid sample generally produce stronger signals than those that have mismatches. Since the sequence and position of each probe immobilized on the chip is known, the identity of the nucleic acid hybridized to a given probe can be determined.

For single-nucleotide polymorphism analyses, sets of four oligonucleotide probes (one for each base type), preferably sets of two oligonucleotide probes (one for each base type of the biallelic marker) are generally designed that span each position of a portion of the candidate region found in the nucleic acid sample, differing only in the identity of the polymorphic base. The relative intensity of hybridization to each series of probes at a particular location allows the identification of the base corresponding to the polymorphic base of the probe. Since biallelic polymorphism detection involves identifying single-base mismatches on the nucleic acid sample, greater hybridization stringencies are required (at lower salt concentration and higher temperature over shorter time periods).

The use of direct electric field control improves the determination of single base mutations (Nanogen). A positive field increases the transport rate of negatively charged nucleic acids and results in a 10-fold increase of the hybridization rates. Using this technique, single base pair mismatches are detected in less than 15 sec (see Sosnowski et al., 1997).

5) Integrated Systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage controls the liquid flow at intersections between the micro-machined channels and changes the liquid flow rate for pumping across different sections of the microchip.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection. In a first step, the DNA samples are amplified, preferably by PCR. Then, the amplification products are subjected to automated microsequencing reactions using

ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide microsequencing primers which hybridize just upstream of the targeted polymorphic base. Once the extension at the 3' end is completed, the primers are separated from the unincorporated fluorescent ddNTPs by capillary electrophoresis. The separation medium used in capillary electrophoresis can for example be polyacrylamide, polyethyleneglycol or dextran. The incorporated ddNTPs in the single-nucleotide primer extension products are identified by fluorescence detection. This microchip can be used to process at least 96 to 384 samples in parallel. It can use the usual four color laser induced fluorescence detection of the ddNTPs.

IX. Association Studies

The identification of genes associated with a particular trait such as asthma susceptibility or individual response to anti-asthmatic drugs can be carried out through two main strategies currently used for genetic mapping: linkage analysis and association studies. Linkage analysis involves the study of families with multiple affected individuals and is now useful in the detection of mono- or oligogenic inherited-traits. Conversely, association studies examine the frequency of marker alleles in unrelated trait positive (T+) individuals compared with control individuals who are randomly selected or preferably trait negative (T-) controls, and are generally employed in the detection of polygenic inheritance.

Association studies as a method of mapping genetic traits rely on the phenomenon of linkage disequilibrium. If two genetic loci lie on the same chromosome, then sets of alleles of these loci on the same chromosomal segment (called haplotypes) tend to be transmitted as a block from generation to generation. When not broken up by recombination, haplotypes can be tracked not only through pedigrees but also through populations. The resulting phenomenon at the population level is that the occurrence of pairs of specific alleles at different loci on the same chromosome is not random, and the deviation from random is called linkage disequilibrium (LD).

If a specific allele in a given gene is directly involved in causing a particular trait T, its frequency will be statistically increased in a T+ population when compared to the frequency in a T- population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele (TCA) will also be increased in T+ individuals compared to T- individuals. Therefore, association between the trait and any allele in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular allele's region. Linkage disequilibrium allows the relative frequencies in T+ and T- populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles.

Two alternative approaches can be employed to perform association studies: a genome-wide association study and a candidate gene association study. The genome-wide association study relies on the screening of genetic markers evenly spaced and covering the entire genome. The candidate

gene approach is based on the study of genetic markers specifically located in genes potentially involved in a biological pathway related to the trait of interest. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available.

The general strategy to perform association studies using biallelic markers derived from a candidate gene is to scan two group of individuals (trait + and trait – control individuals which are characterized by a well defined phenotype as described below) in order to measure and statistically compare the allele frequencies of such biallelic markers in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (the associated allele is the TCA), or the associated allele is in linkage disequilibrium with the TCA. The specific characteristics of the associated allele with respect to the candidate gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the TCA but is in linkage disequilibrium with the real TCA, then the TCA can be found by sequencing the vicinity of the associated marker.

It is another object of the present invention to provide a method for the identification and characterization of an association between alleles for one or several biallelic markers of the sequence of the *FLAP* gene and a trait. The method of detecting an association between a genotype and a trait, comprising the steps of: a) determining the frequency of at least one *FLAP*-related biallelic marker in trait positive population according to a method of the invention; b) determining the frequency of at least one *FLAP*-related biallelic marker in a control population according to a method of the invention; and c) determining whether a statistically significant association exists between said genotype and said trait; Optionally, said biallelic markers are selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally, the trait is either a disease, preferably a disease involving the leukotriene pathway, most preferably asthma, a beneficial response to treatment with agents acting on the leukotriene pathway or side-effects related to treatment with agents acting on the leukotriene pathway. Optionally, said

genotyping steps a) and b) may be performed on a pooled biological sample derived from each of said populations; Optionally, said genotyping steps a) and b) are performed separately on biological samples derived from each individual in said population or a subsample thereof; Optionally, said control individuals are trait negative or random controls.

The invention also encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping at least one *FLAP*-related biallelic marker according to a method of the invention for each individual in said population; b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, particularly in "Statistical methods", or those following, specified alone or in any combination; Optionally, said biallelic markers are selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally, said haplotype determination method is performed by asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark algorithm, or an expectation-maximization algorithm.

The present invention also provides a method for the identification and characterization of an association between a haplotype comprising alleles for several biallelic markers of the genomic sequence of the *FLAP* gene and a trait. The method comprises the steps of: a) genotyping a group of biallelic markers according to the invention in trait positive and control individuals; and b) establishing a statistically significant association between a haplotype and the trait. In a further embodiment, a method for the identification and characterization of an association between a haplotype comprising alleles for several biallelic markers of the genomic sequence of the *FLAP* gene and a trait comprises the steps of: a) estimating the frequency of at least one haplotype in a trait positive population according to a method of the invention; b) estimating the frequency of said haplotype in a control population according to a method of the invention; and c) determining

whether a statistically significant association exists between said haplotype and said trait. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following; Optionally, said biallelic markers are selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally, the trait is a disease, preferably a disease involving the leukotriene pathway, most preferably asthma, a beneficial response to treatment with agents acting on the leukotriene pathway or side-effects related to treatment with agents acting on the leukotriene pathway; Optionally, said control individuals are trait negative or random controls. Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c).

If the trait is a beneficial response or conversely a side-effect to treatment with an agent acting on the leukotriene pathway, the method of the invention referred to above further comprises some or all of the following steps: a) selecting a population or cohort of subjects diagnosed as suffering from a specified disease involving the leukotriene pathway; b) administering a specified agent acting on the leukotriene pathway to said cohort of subjects; c) monitoring the outcome of drug administration and identifying those individuals that are trait positive or trait negative relative to the treatment; d) taking from said cohort biological samples containing DNA and testing this DNA for the presence of a specific allele or of a set of alleles for biallelic markers of the *FLAP* gene; e) analyzing the distribution of alleles for biallelic markers between trait positive and trait negative individuals; and, f) performing a statistical analysis to determine if there is a statistically significant association between the presence or absence of alleles of biallelic markers of the *FLAP* gene and the treatment related trait. Optionally, said biallelic markers are selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic

markers are selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A14 or A16 and the complements thereof. The step of testing for and detecting the presence of DNA comprising specific alleles of a biallelic marker or a group of biallelic markers of the present invention can be carried out as described in the present invention.

The invention also encompasses methods of determining whether an individual is at risk of developing asthma, comprising the steps of: a) genotyping at least one *FLAP*-related biallelic marker according to a method of the present invention; and b) correlating the result of step a) with a risk of developing asthma; optionally wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A28; optionally, wherein said *FLAP*-related biallelic marker is selected from the following list of biallelic markers: A2, A14, A16, A18, A19, A22, and A23; and optionally, wherein said *FLAP*-related biallelic marker is the biallelic marker A19.

1) Collection Of DNA Samples From Trait Positive (Trait +) And Control Individuals (Inclusion Criteria)

Population-based association studies do not concern familial inheritance but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (unaffected or trait negative or random) individuals. Preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is ethnically matched to the case population. Moreover, the control group is preferably matched to the case-population for the main known confusion factor for the trait under study (for example age-matched for an age-dependent trait). Ideally, individuals in the two samples are paired in such a way that they are expected to differ only in their disease status. In the following "trait positive population", "case population" and "affected population" are used interchangeably.

In order to perform efficient and significant association studies such as those described herein, the trait under study should preferably follow a bimodal distribution in the population under study, presenting two clear non-overlapping phenotypes, trait + and trait -.

Nevertheless, even in the absence of such bimodal distribution (as may in fact be the case for more complex genetic traits), any genetic trait may still be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait + and trait – phenotypic groups. The selection procedure involves selecting individuals at opposite ends of the non-bimodal phenotype spectra of the trait under study, so as to include in these trait + and trait – populations individuals which clearly represent extreme, preferably non-overlapping phenotypes.

The definition of the inclusion criteria for the trait + and trait – populations is an important aspect of the present invention.

Typical examples of inclusion criteria include a disease involving the leukotriene pathway such as asthma or the evaluation of liver transaminase levels following treatment with an anti-asthma drug such as Zileuton. From a statistical viewpoint, if one considers that in a given population liver transaminase levels follow a standard distribution curve, individuals with extreme phenotypes according to the optimal inclusion criteria would correspond respectively to those exhibiting the lowest liver transaminase levels and those exhibiting the highest liver transaminase levels.

The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

Generally, trait + and trait – populations to be included in association studies such as those described in the present application consist of phenotypically homogenous populations of individuals each representing 100% of the corresponding trait if the trait distribution is bimodal.

If the trait distribution is non-bimodal, trait + and trait – populations consist of phenotypically uniform populations of individuals representing between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and most preferably between 4 and 35% of the total population under study, and selected from individuals exhibiting the extreme phenotypes of the group. The clearer is the difference between the two trait phenotypes, the greater is the probability to observe an association with biallelic markers.

A first group of between 50 and 300 trait + individuals, preferably about 100 individuals, are recruited according to clinical inclusion criteria based on either 1°) affection by disease(s) involving the leukotriene pathway, preferably asthma, 2°) evidence of side-effects observed following administration of an agent acting on the leukotriene pathway, preferably increased liver transaminase levels following administration of Zileuton, or 3°) evidence of particular responses to treatment with agents acting on the leukotriene pathway.

In each case, a similar number of trait negative individuals are included in such studies. They are checked for the absence of the clinical criteria defined above. Both trait + and trait - individuals should be unrelated cases.

In the context of the present invention, one association study were carried out. The considered trait was asthma. Collection of DNA samples from trait + and trait – individuals is described in Example 5.

2) Genotyping Of Trait + And Trait - Individuals

Allelic frequencies of the biallelic markers in each of the above described populations can be determined using one of the methods described above under the heading " Methods of Genotyping DNA samples for Biallelic Markers ". Analyses are preferably performed on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual in similar conditions as those described above for the generation of biallelic markers.

In a preferred embodiment, amplified DNA samples are subjected to automated microsequencing reactions using fluorescent ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide microsequencing primers which hybridize just upstream of the polymorphic base. Genotyping is further described in Example 5.

3) Single marker Association Studies and Haplotype frequency analysis

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and control populations, and comparing these frequencies with a statistical test to determine if their is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype may be used. Preferably the statistical test employed is a chi-square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about 1×10^{-2} or less, more preferably about 1×10^{-4} or less, for a single biallelic marker analysis and about 1×10^{-3} or less, still more preferably 1×10^{-6} or less and most preferably of about 1×10^{-8} or less, for a haplotype analysis involving two or more markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and a disease involving the leukotriene pathway can be revealed and used for diagnosis and drug screening purposes.

To address the problem of false positives similar analysis may be performed with the same case-control populations in random genomic regions. Results in random regions and the candidate region are compared as described in a co-pending US Provisional Patent Application entitled "Methods, Software And Apparati For Identifying Genomic Regions Harboring A Gene Associated

With A Detectable Trait," U.S. Serial Number 60/107,986, filed November 10, 1998, the contents of which are incorporated herein by reference.

Single marker association

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers, usually between 2 and 10 markers, is determined in the trait + and trait - populations. In a second phase of the analysis, the position of the genetic loci responsible for the given trait is further refined using a higher density set of markers. However, if the candidate gene under study is relatively small in length, as it is the case for the *FLAP* gene, it is believed that a single phase is sufficient to establish significant associations.

In one preferred embodiment of the invention in which a correlation was found between a set of biallelic markers of the *FLAP* gene and a disease involving the leukotriene pathway, more particularly asthma, results of the first step of the association study, further details of which are provided in example 7, seem to indicate that asthma is associated most strongly with the biallelic marker A19 (10-35/390, allele T). Further details concerning these associations are provided in Example 7.

Similar association studies can also be carried out with other biallelic markers within the scope of the invention, preferably with biallelic markers in linkage disequilibrium with the markers associated with asthma, including the biallelic markers A1 to A28.

Similar associations studies can be routinely carried out by the skilled technician using the biallelic markers of the invention which are defined above with different trait + and trait - populations. Suitable further examples of possible association studies using biallelic markers of the *FLAP* gene, including the biallelic markers A1 to A28, involve studies on the following populations:

- a trait + population suffering from a disease involving the leukotriene pathway and a healthy unaffected population; or

- a trait + population treated with agents acting on the leukotriene pathway suffering from side-effects resulting from the treatment and an trait - population treated with same agents without any side-effects; or

- a trait + population treated with agents acting on the leukotriene pathway showing a beneficial response and a trait – population treated with same agents without any beneficial response.

Haplotype frequency analysis

A haplotype analysis is interesting in that it increases the statistical significance of an analysis involving individual markers. Indeed, by combining the informativeness of a set of biallelic markers, it increases the value of the results obtained through association analyses, allowing false positive and/or negative data that may result from the single marker studies to be eliminated.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined and compared for distinct populations of trait + and trait - individuals. The number of trait + individuals which should be subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300 , with a preferred number of individuals ranging between 50 and 150. The same considerations apply to the number of unaffected controls used in the study.

The results of this first analysis provide haplotype frequencies for the tested trait + and trait - individuals, and the estimated p value for each evaluated haplotype.

In the association of the biallelic markers of *FLAP* gene with the asthma, several haplotypes were also shown to be significant (see Figure 3). For example, the preferred haplotypes comprise the allele T of the biallelic marker A19 (10-35/390). The more preferred haplotype (HAP 1 of Figure 3) comprise the allele A of the marker A14 (10-33/234) and the allele T of the marker A19 (10-35/390). This haplotype is considered to be highly significant of an association with asthma. The other significant haplotypes are detailed in Example 8.

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control individuals are pooled and randomized with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups, which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is reiterated preferably at least between 100 and 10000 times. The repeated iterations allow the determination of the probability to obtain by chance the tested haplotype.

For the association between asthma and the three considered haplotypes, a randomized haplotype analysis was reiterated 1000 times or 10000 times and the results are shown in Figure 4. These results demonstrate that among 1000 iterations none and among 10,000 iterations only 1 of the obtained haplotypes had a p-value comparable to the one obtained for the haplotype HAP1. These results clearly validate the statistical significance of the association between this haplotype and asthma.

Using the method described above and evaluating the associations for single marker alleles or for haplotypes permits an estimation of the risk a corresponding carrier has to develop a given trait, and particularly in the context of the present invention, a disease, preferably a disease involving the leukotriene pathway, more preferably asthma. Significance thresholds of relative risks are to be adapted to the reference sample population used. The evaluation of the risk factors is detailed in "Statistical methods".

It will of course be understood by practitioners skilled in the treatment of diseases involving the leukotriene pathway listed above, and in particular asthma, that the present invention does not

intend to provide an absolute identification of individuals who could be at risk of developing a particular disease involving the leukotriene pathway or who will or will not respond or exhibit side-effects to treatment with agents acting on the leukotriene pathway but rather to indicate a certain degree or likelihood of developing a disease or of observing in a given individual a response or a side-effect to treatment with said agents.

However, this information is extremely valuable as it can, in certain circumstances, be used to initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms. In diseases such as asthma, in which attacks may be extremely violent and sometimes fatal if not treated on time, the knowledge of a potential predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy. Similarly, a diagnosed predisposition to a potential side-effect could immediately direct the physician toward a treatment for which such side-effects have not been observed during clinical trials.

X. Statistical Methods

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation may be used.

1) Methods To Estimate Haplotype Frequencies In A Population

The gametic phase of haplotypes is unknown when diploid individuals are heterozygous at more than one locus. Using genealogical information in families gametic phase can sometimes be inferred (Perlin et al., 1994). When no genealogical information is available different strategies may be used. One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR amplification (see Newton et al, 1989; Wu et al., 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano et al., 1990). Further, a sample may be haplotyped for sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S. S., 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalization at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm to infer the phase of PCR-amplified DNA genotypes introduced by Clark, A.G.(1990) may be used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognized haplotypes. For each positive identification, the complementary haplotype is added to the list of recognized haplotypes, until the phase information

for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Alternatively, one can use methods estimating haplotype frequencies in a population without assigning haplotypes to each individual. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster et al., 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used (see Excoffier L. and Slatkin M., 1995). The EM algorithm is a generalized iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. The EM algorithm can be applied using for example the EM-HAPLO program (Hawley M. E. et al., 1994) or the Arlequin program (Schneider et al., 1997). Any other method known in the art to determine or to estimate the frequency of a haplotype in a population may be used(see Lange K., 1997; Weir, B.S., 1996). The EM algorithm is briefly described below.

A sample of N unrelated individuals is typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can categorized in F different phenotypes. Suppose that we have H underlying possible haplotypes (in case of K biallelic markers, $H=2^K$).

For phenotype j, suppose that c_j genotypes are possible. We thus have the following equation

$$P_j = \sum_{i=1}^{c_j} pr(genotype_i) = \sum_{i=1}^{c_j} pr(h_k, h_l) \qquad \text{Equation 1}$$

where Pj is the probability of the phenotype j, h_k and h_l are the two haplotypes constituent the genotype i. Under the Hardy-Weinberg equilibrium, $pr(h_k h_l)$ becomes:

$$pr(h_k, h_l) = pr(h_k)^2 \text{ if } h_k = h_l,\ pr(h_k, h_l) = 2\,pr(h_k).pr(h_l) \text{ if } h_k \neq h_l. \quad \text{Equation 2}$$

The successive steps of the E-M algorithm can be described as follows:

Starting with initial values of the of haplotypes frequencies, noted $p_1^{(0)}, p_2^{(0)}, \ldots p_H^{(0)}$, these initial values serve to estimate the genotype frequencies (Expectation step) and then estimate another set of haplotype frequencies (Maximization step), noted $p_1^{(1)}, p_2^{(1)}, \ldots p_H^{(1)}$, these two steps are iterated until changes in the sets of haplotypes frequency are very small.

A stop criterion can be that the maximum difference between haplotype frequencies between two iterations is less than 10^{-7}. These values can be adjusted according to the desired precision of estimations.

At a given iteration s, the Expectation step consists in calculating the genotypes frequencies by the following equation:

$$pr(genotype_i)^{(s)} = pr(phenotype_j).pr(genotype_i|phenotype_j)^{(s)}$$

$$= \frac{n_j}{N} \cdot \frac{pr(h_k,h_l)^{(s)}}{p_j^{(s)}} \qquad \text{Equation 3}$$

where genotype i occurs in phenotype j, and where h_k and h_l constitute genotype i. Each probability is derived according to eq. 1, and eq. 2 described above.

Then the Maximization step simply estimates another set of haplotype frequencies given the genotypes frequencies. This approach is also known as the gene-counting method (Smith, 1957).

$$p_t^{(s+1)} = \frac{1}{2} \sum_{j=1}^{F} \sum_{i=1}^{c_j} \delta_{it} . pr(genotype_i)^{(s)} \qquad \text{Equation 4}$$

Where δ_{it} is an indicator variable which count the number of time haplotype t in genotype i. It takes the values of 0, 1 or 2.

To ensure that the estimation finally obtained is the maximum-likelihood estimation several values of departures are required. The estimations obtained are compared and if they are different the estimations leading to the best likelihood are kept.

2) Methods To Calculate Linkage Disequilibrium Between Markers

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_i, M_j) having alleles (a_i/b_i) at marker M_i and alleles (a_j/b_j) at marker M_j can be calculated for every allele combination (a_i,a_j; a_i,b_j; b_i,a_j and b_i,b_j), according to the Piazza formula:

$\Delta_{aiaj} = \sqrt{\theta 4} - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)}$, where:

$\theta 4$= - - = frequency of genotypes not having allele a_i at M_i and not having allele a_j at M_j

$\theta 3$= - + = frequency of genotypes not having allele a_i at M_i and having allele a_j at M_j

$\theta 2$= + - = frequency of genotypes having allele a_i at M_i and not having allele a_j at M_j

Linkage disequilibrium (LD) between pairs of biallelic markers (M_i, M_j) can also be calculated for every allele combination (ai,aj; ai,bj; b_i,a_j and b_i,b_j), according to the maximum-likelihood estimate (MLE) for delta (the composite genotypic disequilibrium coefficient), as described by Weir (Weir B. S., 1996). The MLE for the composite linkage disequilibrium is:

$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4/2)/N - 2(pr(a_i).\ pr(a_j))$

Where $n_1 = \Sigma$ phenotype (a_i/a_i, a_j/a_j), $n_2 = \Sigma$ phenotype (a_i/a_i, a_j/b_j), $n_3 = \Sigma$ phenotype (a_i/b_i, a_j/a_j), n4= Σ phenotype (a_i/b_i, a_j/b_j) and N is the number of individuals in the sample.

This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, M_i (a_i/b_i) and M_j (a_j/b_j), fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between *ai* and *aj* is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where $pr(a_i)$ is the probability of allele a_i and $pr(a_j)$ is the probability of allele a_j and where $pr(haplotype\ (a_i, a_j))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between M_i and M_j.

Then a normalized value of the above is calculated as follows:

$$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i).\ pr(a_j), -pr(b_i).\ pr(b_j)) \text{ with } D_{aiaj}<0$$

$$D'_{aiaj} = D_{aiaj} / \max(pr(b_i).\ pr(a_j), pr(a_i).\ pr(b_j)) \quad \text{with } D_{aiaj}>0$$

The skilled person will readily appreciate that other linkage disequilibrium calculation methods can be used.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 75 and 200, more preferably around 100.

3) Evaluation Of Risk Factors

The association between a risk factor (in genetic epidemiology the risk factor is the presence or the absence of a certain allele or haplotype at marker loci) and a disease is measured by the odds ratio (OR) and by the relative risk (RR). If $P(R^+)$ is the probability of developing the disease for individuals with R and $P(R^-)$ is the probability for individuals without the risk factor, then the relative risk is simply the ratio of the two probabilities, that is:

$$RR = P(R^+)/P(R^-)$$

In case-control studies, direct measures of the relative risk cannot be obtained because of the sampling design. However, the odds ratio allows a good approximation of the relative risk for low-incidence diseases and can be calculated:

$$OR = (F^+/(1-F^+))/(F^-/(1-F^-))$$

F^+ is the frequency of the exposure to the risk factor in cases and F^- is the frequency of the exposure to the risk factor in controls. F^+ and F^- are calculated using the allelic or haplotype frequencies of the study and further depend on the underlying genetic model (dominant, recessive, additive...).

One can further estimate the attributable risk (AR) which describes the proportion of individuals in a population exhibiting a trait due to a given risk factor. This measure is important in

quantifying the role of a specific factor in disease etiology and in terms of the public health impact of a risk factor. The public health relevance of this measure lies in estimating the proportion of cases of disease in the population that could be prevented if the exposure of interest were absent. AR is determined as follows:

$AR = P_E (RR-1) / (P_E (RR-1)+1)$

AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype. P_E is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which, is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

XI. Identification Of Biallelic Markers In Linkage Disequilibrium With The Biallelic Markers Of The Present Invention.

Once a first biallelic marker has been identified in a genomic region of interest, the practitioner of ordinary skill in the art, using the teachings of the present invention, can easily identify additional biallelic markers in linkage disequilibrium with this first marker. As mentioned before any marker in linkage disequilibrium with a first marker associated with a trait will be associated with the trait. Therefore, once an association has been demonstrated between a given biallelic marker and a trait, the discovery of additional biallelic markers associated with this trait is of great interest in order to increase the density of biallelic markers in this particular region. The causal gene or mutation will be found in the vicinity of the marker or set of markers showing the highest correlation with the trait.

The invention also concerns a method for the identification and characterization of a biallelic marker in linkage disequilibrium with a biallelic marker of a *FLAP* gene, preferably a biallelic marker of a *FLAP* gene of which one allele is associated with a trait. In one embodiment, the biallelic marker in linkage disequilibrium with a biallelic marker of the *FLAP* gene is in the genomic region harboring the *FLAP* gene, but outside of the *FLAP* gene itself. In another embodiment, the biallelic marker in linkage disequilibrium with a biallelic marker of the *FLAP* gene is itself located within the *FLAP* gene. The method comprises the following steps: a) amplifying a genomic fragment comprising a first biallelic marker from a plurality of individuals; b) identifying second biallelic markers in the genomic region harboring the first biallelic marker; c) conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and d) selecting said second biallelic markers in linkage disequilibrium with said first marker.

In one embodiment, the step of sequencing and identifying second biallelic markers comprises sequencing second biallelic markers within the *FLAP* gene. In a further embodiment, the step of sequencing and identifying second biallelic markers comprises sequencing second biallelic markers within the amplified region of the *FLAP* gene.

Once identified, the sequences in linkage disequilibrium with a biallelic marker of the *FLAP* gene may be used in any of the methods described herein, including methods for determining an

association between biallelic marker and a trait, methods for identifying individuals having a predisposition for a trait, methods of disease treatment, methods of identifying individuals likely to respond positively or negatively to drug treatment, and methods of using drugs. In particular, biallelic markers in linkage disequilibrium with a biallelic marker in the *FLAP* gene may be used to identify individuals having a predisposition to asthma or to positive or negative responses to treatment with anti-asthma drugs such as Zileuton.

Methods to identify biallelic markers and to conduct linkage disequilibrium analysis are described herein in "Statistical methods" and can be carried out by the skilled person without undue experimentation. The present invention then also concerns biallelic markers which are in linkage disequilibrium with the specific biallelic markers A1 to A28 and which are expected to present similar characteristics in terms of their respective association with a given trait.

XII. Identification Of Trait-Causing Mutations In The *FLAP* Gene

Mutations in the *FLAP* gene which are responsible for a detectable phenotype may be identified by comparing the sequences of the *FLAP* genes from trait-positive and trait-negative individuals. Preferably, trait + individuals to be sequenced carry the haplotype shown to be associated to the trait and trait – individuals to be sequenced do not carry the haplotype associated to the trait. The detectable phenotype may comprise a variety of manifestations of altered FLAP function, including a disease involving the leukotriene pathway, a response to an agent acting on the leukotriene pathway or side-effects linked to a treatment with this agent. The mutations may comprise point mutations, deletions, or insertions in the *FLAP* gene. The mutations may lie within the coding sequence for the FLAP protein or within regulatory regions in the *FLAP* gene.

The method used to detect such mutations generally comprises the following steps: a) amplification of a region of the *FLAP* gene comprising a biallelic marker or a group of biallelic markers associated with the trait from DNA samples of trait positive patients and trait negative controls; b) sequencing of the amplified region; c) comparison of DNA sequences from trait-positive patients and trait-negative controls; and, d) determination of mutations specific to trait-positive patients.

Oligonucleotide primers are constructed as described previously to amplify the sequences of each of the exons, introns or the promoter region of the *FLAP* gene.

Each primer pair is used to amplify the exon or promoter region from which it is derived. Amplification is carried out on genomic DNA samples from trait positive patients and trait negative controls, preferably using the PCR conditions described in the examples. Amplification products from the genomic PCRs are then subjected to sequencing, preferably through automated dideoxy terminator sequencing reactions and electrophoresed, preferably on ABI 377 sequencers. Following gel image analysis and DNA sequence extraction, ABI sequence data are automatically analyzed to detect the presence of sequence variations among trait positive and trait negative individuals. Sequences are verified by determining the sequences of both DNA strands for each individual.

Candidate polymorphisms suspected of being responsible for the detectable phenotype, such as a disease, a beneficial response to an agent acting on the leukotriene pathway or side-effects linked to a treatment with this agent, are then verified by screening a larger population of trait positive and trait negative individuals using polymorphism analysis techniques such as the techniques described above. Polymorphisms which exhibit a statistically significant correlation with the detectable phenotype are deemed responsible for the detectable phenotype.

Most of the biallelic polymorphisms of the *FLAP* gene observed in the context of the present invention do not appear to drastically modify the amino acid sequence of the FLAP protein. Also, they do not seem to be located in splicing sequences. However, they may be associated with changes in basic *FLAP* expression in one or more tissues. Such polymorphisms may eventually modify the transcription rate of *FLAP* DNA, *FLAP* mRNA stability, or the translation rate of *FLAP* mRNA.

The biallelic polymorphisms may also be associated with changes in the modulation of *FLAP* expression through expression modifiers. The term "expression modifier" is intended to encompass chemical agents that modulate the action of FLAP through modulation of *FLAP* gene expression.

The basic *FLAP* expression levels in different tissues can be determined by analyses of tissue samples from individuals typed for the presence or absence of a specific polymorphism. Any convenient method can be used such as ELISA, RIA for protein quantitation, and such as Nothern blot or other hybridization analyses, and quantitative RT-PCR for mRNA quantitation. The tissue specific expression can then be correlated with the genotype. More details on some of these methods are provided below under the heading "Screening of agents".

Furthermore, the strong association observed for the first time between the *FLAP* gene and asthma confirms the need to locate and study any mutation of the *FLAP* gene as such mutation is susceptible of having an incidence on leukotriene metabolism and hence on the therapeutic choices made when considering various treatment alternatives for an individual with a particular condition involving the leukotriene pathway.

There are numerous possibilities for causal mutations within the *FLAP* gene. One of the causal mutations can be an amino acid change in the FLAP protein which can lead to alterations in FLAP substrate specificity and/or activity. Methods for analyzing protein-protein or protein-ligand interactions are detailed below under the heading "Screening of agents".

Another possible causal mutation of the *FLAP* gene is a modification in its regulatory region, and particularly in the sequence of its native promoter. This type of mutation can be studied through the determination of basic expression levels by expression assays for the particular promoter sequence. The assays may be performed with the FLAP coding sequence or with a detectable marker sequence. To determine tissue specificity, the assay is performed in cells from different sources. Some methods are discussed in more detail below under the heading "Screening of agents".

When used herein, the term "basic expression levels" intends to designate *FLAP* expression levels normally observed in individuals not bearing the associated allele of biallelic markers of the present invention.

In another embodiment, the mutant *FLAP* allele which causes a detectable phenotype can be isolated by obtaining a nucleic acid sample such as a genomic library or a cDNA library from an individual expressing the detectable phenotype. The nucleic acid sample can be contacted with one or more probes lying in the region of the *FLAP* gene where the associated biallelic marker or group of biallelic markers or with PCR-typeable primers specific to the amplification of this biallelic marker or group of biallelic markers. The mutation can be identified by conducting sequencing reactions on the nucleic acids which hybridize with the probes defined herein or which show amplification by PCR. The region of the *FLAP* gene containing the mutation responsible for the detectable phenotype may be used in diagnostic techniques such as those described below. For example, microsequencing oligonucleotides, or oligonucleotides containing the mutation responsible for the detectable phenotype for amplification, or hybridization based diagnostics, such as those described herein, may be used for detecting individuals suffering from the detectable phenotype or individuals at risk of developing the detectable phenotype at a subsequent time. In addition, the *FLAP* allele responsible for the detectable phenotype may be used in gene therapy. The *FLAP* allele responsible for the detectable phenotype may also be cloned into an expression vector to express the mutant FLAP protein as described herein.

XIII. Biallelic Markers Of The Invention In Methods Of Genetic Diagnostics

The biallelic markers of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time. The trait analyzed using the present diagnostics may be any detectable trait, including a disease involving the leukotriene pathway, a beneficial response to treatment with agents acting on the leukotriene pathway or side-effects related to treatment with agents acting on the leukotriene pathway.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

The present invention provides diagnostic methods to determine whether an individual is at risk of developing a disease or suffers from a disease resulting from a mutation or a polymorphism in the *FLAP* gene. The present invention also provides methods to determine whether an individual is likely to respond positively to an agent acting on the leukotriene pathway or whether an individual is at risk of developing an adverse side-effect to an agent acting on the leukotriene pathway.

These methods involve obtaining a nucleic acid sample from the individual and, determining, whether the nucleic acid sample contains at least one allele or at least one biallelic marker haplotype, indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular *FLAP* polymorphism or mutation (trait-causing allele).

Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described above in VIII. The diagnostics may be based on a single biallelic marker or a on group of biallelic markers.

In each of these methods, a nucleic acid sample is obtained from the test subject and the biallelic marker pattern of one or more of the biallelic markers A1 to A28, the complements thereof or a biallelic marker in linkage disequilibrium therewith is determined.

In one embodiment, a PCR amplification is conducted on the nucleic acid sample to amplify regions in which polymorphisms associated with a detectable phenotype have been identified. The amplification products are sequenced to determine whether the individual possesses one or more *FLAP* polymorphisms associated with a detectable phenotype. The primers used to generate amplification products may comprise the primers B1 to B17 and C1 to C17. Alternatively, the nucleic acid sample is subjected to microsequencing reactions as described above to determine whether the individual possesses one or more *FLAP* polymorphisms associated with a detectable phenotype resulting from a mutation or a polymorphism in the *FLAP* gene. The primers used in the microsequencing reactions may include the primers D1 to D28 and E1 to E28. In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotide probes which, specifically hybridize to one or more *FLAP* alleles associated with a detectable phenotype. The probes used in the hybridization assay may include the probes P1 to P28, a complementary sequence thereto or a fragment thereof comprising the polymorphic base. In another embodiment, the nucleic acid sample is contacted with a second *FLAP* oligonucleotide capable of producing an amplification product when used with the allele specific oligonucleotide in an amplification reaction. The presence of an amplification product in the amplification reaction indicates that the individual possesses one or more *FLAP* alleles associated with a detectable phenotype.

In a preferred embodiment, the identity of the nucleotide present at, at least one biallelic marker selected from the group consisting of A2, A14, A16, A18, A19, A22, and A23, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith, is determined and the detectable trait is asthma. In another preferred embodiment the identity of the nucleotide present at, at least one of the polymorphic sites selected from the group consisting of A14 and A19, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith, is determined. In more preferred embodiment, the identity of the nucleotide present at the polymorphic site A19, and the complements thereof, or optionally the biallelic markers in linkage

disequilibrium therewith, is determined. Diagnostic kits comprising polynucleotides of the present invention are further described in the present invention.

These diagnostic methods are extremely valuable as they can, in certain circumstances, be used to initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms. In diseases in which attacks may be extremely violent and sometimes fatal if not treated on time, such as asthma, the knowledge of a potential predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at a *FLAP*-related biallelic marker. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or a mismatch detection assay based on polymerases and/or ligases. The diagnostic kits can be manufactured to perform any of the genotyping methods described in the current application using manufacturing and formulation methods commonly in the art. Preferably such a kit may provide for the determination of the allele of a biallelic marker selected from *FLAP*-related biallelic markers. Optionally such a kit may include instructions for scoring the results of the determination with respect to the test subjects' risk of contracting a disease involving the leukotriene pathway, a beneficial response to treatment with agents acting on the leukotriene pathway or side-effects related to treatment with agents acting on the leukotriene pathway.

XIV. Treatment Of Diseases Involving The Leukotriene Pathway

The invention also relates to a method of determining whether a subject is likely to respond positively to treatment with a medicament, preferably a medicament acting directly or indirectly on the leukotriene pathway.

The method comprises identifying a first population of individuals who response positively to said medicament and a second population of individuals who respond negatively to said medicament. One or more biallelic markers is identified in the first population which is associated with a positive response to said medicament or one or more biallelic markers is identified in the second population which is associated with a negative response to said medicament. The biallelic markers may be identified using the techniques described herein.

The DNA sample is then obtained form the subject tested. The DNA sample is analyzed to determine whether it comprises one or more alleles of biallelic markers associated with a positive response to a medicament or one or more alleles of biallelic markers associated with a negative response to treatment with the medicament. In some embodiments, the DNA sample is analyzed to

identify subjects whose DNA comprises one or more alleles of biallelic markers associated with a positive response to the medicament and whose DNA lacks one or more alleles of biallelic markers associated with a negative response to treatment with the medicament.

In other embodiments, the medicament is administered to the subject in a clinical trial if the DNA sample contains one or more alleles of biallelic markers associated with positive response to the medicament and/or if the DNA sample lacks one or more alleles of biallelic markers associated with a negative response to treatment with the medicament. In preferred embodiments, the medicament is an anti-asthma drug such as Zileuton. In other embodiments, the negative response comprises one or more side-effects, such as increased liver transaminase levels. Using the methods of the present invention, the evaluation of drug efficacy may be conducted in a population of individuals likely to respond favorably to the medicament.

The invention also concerns a method for the clinical testing of a medicament, preferably a medicament acting directly or indirectly on the leukotriene pathway. The method comprises the following steps: a) administering a medicament, preferably a medicament capable of acting directly or indirectly on the leukotriene pathway to a heterogeneous population of individuals; b) identifying a first population of individuals who response positively to said medicament and a second population of individuals who respond negatively to said medicament; c) identifying biallelic markers in said first population which are associated with a positive response to said medicament and/or biallelic markers in said second population which are associated with a negative response to said medicament; d) selecting individuals whose DNA comprises one or more alleles of biallelic markers associated with a positive response to said medicament and/or whose DNA lacks one or more alleles of biallelic markers associated with a negative response to said medicament; and, d) administering said medicament to said individuals.

Such methods are deemed to be extremely useful to increase the benefit/risk ratio resulting from the administration of medicaments which may cause undesirable side-effects and/or be inefficacious to a portion of the patient population to which it is normally administered.

Once an individual has been diagnosed as suffering from a disease involving the leukotriene pathway such as asthma, selection tests are carried out to determine whether the DNA of this individual comprises alleles of a biallelic marker or of a group of biallelic markers associated a positive response to treatment or with a negative response to treatment which may include either side-effects or unresponsiveness.

The selection of the patient to be treated using the method of the present invention can be carried out through the detection methods described above. The individuals which are to be selected are preferably those whose DNA does not comprise alleles of a biallelic marker or of a group of biallelic markers associated with negative response to treatment.

Once the patient's genetic predispositions have been determined, the clinician can select appropriate treatment for which the particular side-effect observed for the patient has not been

reported or has been reported only marginally and preferably from an allelic association which does not involve the same biallelic marker or markers as those found in the DNA of the patient. Several drugs useful in the treatment of diseases involving the leukotriene pathway may be chosen. Compounds acting on the leukotriene pathway are described for example in US patents 4,873,259; 4,970,215; 5,310,744; 5,225,421; and 5,081,138, or in EP 0 419 049, the disclosures of which are incorporated by reference.

XV. FLAP Proteins And Polypeptide Fragments

The term "FLAP polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies FLAP proteins from humans, including isolated or purified FLAP proteins consisting, consisting essentially, or comprising the sequence of SEQ ID No 3 and comprising an isoleucine at position 127 in SEQ ID No 3. It should be noted the FLAP proteins of the invention are based on the naturally-occurring variant of the amino acid sequence of human FLAP, wherein the valine residue of amino acid position 127 in SEQ ID No 3 has been replaced with an isoleucine residue. This variant protein and the fragments thereof which contain amino acid position 127 of SEQ ID No 3 are collectively referred to herein as "127-Ile variants" or 127-Ile FLAP polypeptides".

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 in SEQ ID No 3. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the FLAP protein sequence.

FLAP proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes. The FLAP polypeptides of the invention can be made using routine expression methods known in the art. The polynucleotide encoding the desired polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides, and a summary of some of the more common systems. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

In addition, shorter protein fragments is produced by chemical synthesis. Alternatively the proteins of the invention is extracted from cells or tissues of humans or non-human animals.

Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Any *FLAP* cDNA, including SEQ ID No 2, is used to express FLAP proteins and polypeptides. The preferred *FLAP* cDNA comprises the allele A of the biallelic marker A21. The nucleic acid encoding the FLAP protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The *FLAP* insert in the expression vector may comprise the full coding sequence for the FLAP protein or a portion thereof. For example, the *FLAP* derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the FLAP protein of SEQ ID No 3, where in said consecutive amino acids comprising an isoleucine residue in amino acid position 127.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No 5,082,767.

In one embodiment, the entire coding sequence of the *FLAP* cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the FLAP protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the *FLAP* cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the FLAP protein or a portion thereof is obtained by PCR from a bacterial vector containing the *FLAP* cDNA of SEQ ID No 3 using oligonucleotide primers complementary to the *FLAP* cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the FLAP protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

Alternatively, the nucleic acids encoding the FLAP protein or a portion thereof is cloned into pED6dpc2 (Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs is transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded.

The above procedures may also be used to express a mutant FLAP protein responsible for a detectable phenotype or a portion thereof.

The expressed proteins is purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed FLAP protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the FLAP protein or portion thereof is attached to the chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the FLAP protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the FLAP protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the FLAP protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the FLAP protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed FLAP protein or a portion thereof are described below.

If antibody production is not possible, the nucleic acids encoding the FLAP protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the FLAP protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the β-globin gene or the nickel binding polypeptide and

the FLAP protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

Antibodies That Bind FLAP Polypeptides of the Invention

Any FLAP polypeptide or whole protein may be used to generate antibodies capable of specifically binding to expressed FLAP protein or fragments thereof as described. The antibody compositions of the invention are capable of specifically binding or specifically bind to the 127-Ile variant of the FLAP protein. For an antibody composition to specifically bind to the 127-Ile variant of FLAP it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for full length 127-Ile variant of FLAP than for full length 127-Val variant of FLAP in an ELISA, RIA, or other antibody-based binding assay.

In a preferred embodiment of the invention antibody compositions are capable of selectively binding, or selectively bind to an epitope-containing fragment of a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, wherein said epitope comprises an isoleucine residue at amino acid position 127 in SEQ ID No 3 , wherein said antibody composition is optionally either polyclonal or monoclonal.

The present invention also contemplates the use of polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of a FLAP polypeptide in the manufacture of antibodies, wherein said contiguous span comprises an isoleucine residue at amino acid position 127 of SEQ ID No 3. In a preferred embodiment such polypeptides are useful in the manufacture of antibodies to detect the presence and absence of the 127-Ile variant.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of FLAP than the one to which antibody binding is desired, and animals which do not express *FLAP* (i.e. a *FLAP* knock out animal as described in herein) are particularly useful for preparing antibodies. *FLAP* knock out animals will recognize all or most of the exposed regions of FLAP as foreign antigens, and therefore produce antibodies with a wider array of FLAP epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to the 127-Ile variant. In addition, the humoral immune system of animals which produce a

species of FLAP that resembles the antigenic sequence will preferentially recognize the differences between the animal's native FLAP species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to the 127-Ile variant.

XVI. Recombinant Vectors, Cell Hosts, and Transgenic Animals

Recombinant Vectors

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The present invention encompasses a family of recombinant vectors that comprise a regulatory polynucleotide derived from the *FLAP* genomic sequence, or a coding polynucleotide from the *FLAP* genomic sequence. Consequently, the present invention further deals with a recombinant vector comprising either a regulatory polynucleotide comprised in the nucleic acid of SEQ ID No 1 or a polynucleotide comprising the *FLAP* coding sequence or both.

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences and coding sequences, as well as any *FLAP* primer or probe as defined above.

In a first preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from a *FLAP* genomic sequence of SEQ ID No 1 or a *FLAP* cDNA, for example the cDNA of SEQ ID No 2 in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

A second preferred embodiment of the recombinant vectors according to the invention consists of expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express the FLAP polypeptide which can be then purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against the FLAP protein. In other embodiments, the expression vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding a FLAP protein, preferably the FLAP protein of the amino acid sequence of SEQ ID

promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983; O'Reilly et al., 1992), the lambda PR promoter or also the trc promoter.

Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a promoter is well within the ability of a person skilled in the field of genetic egineering. For example, one may refer to the book of Sambrook et al.(1989) or also to the procedures described by Fuller et al.(1996).

Other regulatory elements

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

The vector containing the appropriate DNA sequence as described above, more preferably *FLAP* gene regulatory polynucleotide, a polynucleotide encoding the FLAP polypeptide selected from the group consisting of SEQ ID No 1 or a fragment or a variant thereof and SEQ ID No 2, or both of them, can be utilized to transform an appropriate host to allow the expression of the desired polypeptide or polynucleotide.

3. Selectable Markers

Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression construct. The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

4. Preferred Vectors.

Bacterial vectors

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of other suitable vectors are known to those of skill in the art, and commercially available, such as the following bacterial vectors: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A

(Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

Bacteriophage vectors

The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb.

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994). Recombinant P1 clones comprising *FLAP* nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (Linton et al., 1993). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al.(1994).

Baculovirus vectors

A suitable vector for the expression of the FLAP polypeptide of SEQ ID No 6 or fragments or variants thereof is a baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharmingen) that is used to transfect the SF9 cell line (ATCC N°CRL 1711) which is derived from *Spodoptera frugiperda*.

Other suitable vectors for the expression of the FLAP polypeptide of SEQ ID No 6 or fragments or variants thereof in a baculovirus expression system include those described by Chai et al.(1993), Vlasak et al.(1983) and Lenhard et al.(1996).

Viral vectors

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996) or Ohno et al.(1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application N° FR-93. 05954).

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides *in vivo*, particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include Bryan

high titer (ATCC Nos. VR-334, VR-657, VR-726, VR-659 and VR-728). Other preferred retroviral vectors are those described in Roth et al.(1996), PCT Application No WO 93/25234, PCT Application No WO 94/ 06920, Roux et al., 1989, Julan et al., 1992 and Neda et al., 1991.

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

BAC vectors

The bacterial artificial chromosome (BAC) cloning system (Shizuya et al., 1992) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli*. A preferred BAC vector consists of pBeloBAC11 vector that has been described by Kim et al.(1996). BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

5. Delivery Of The Recombinant Vectors

In order to effect expression of the polynucleotides and polynucleotide constructs of the invention, these constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states.

One mechanism is viral infection where the expression construct is encapsulated in an infectious viral particle.

Several non-viral methods for the transfer of polynucleotides into cultured mammalian cells are also contemplated by the present invention, and include, without being limited to, calcium phosphate precipitation (Graham et al., 1973; Chen et al., 1987;), DEAE-dextran (Gopal, 1985), electroporation (Tur-Kaspa et al., 1986; Potter et al., 1984), direct microinjection (Harland et al.,

1985), DNA-loaded liposomes (Nicolau et al., 1982; Fraley et al., 1979), and receptor-mediate transfection (Wu and Wu, 1987; 1988). Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application N° WO 90/11092 (Vical Inc.) and also in PCT application No WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson et al.(1996) and of Huygen et al.(1996).

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be proceeded with a particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein et al.(1987).

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, 1991; Wong et al., 1980; Nicolau et al., 1987)

In a specific embodiment, the invention provides a composition for the *in vivo* production of the FLAP protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced *in vitro* in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been

transformed with the vector coding for the desired FLAP polypeptide or the desired fragment thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

Cell Hosts

Another object of the invention consists of a host cell that have been transformed or transfected with one of the polynucleotides described therein. Are included host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above.

Generally, a recombinant host cell of the invention comprises any one of the polynucleotides or the recombinant vectors described therein.

A further recombinant cell host according to the invention comprises a polynucleotide containing a biallelic marker selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally, said biallelic marker is selected from the group consisting of A1 to A13, A15, A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith

Preferred host cells used as recipients for the expression vectors of the invention are the following:

a) Prokaryotic host cells: *Escherichia coli* strains (I.E.DH5-α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*.

b) Eukaryotic host cells: HeLa cells (ATCC N°CCL2; N°CCL2. 1; N°CCL2. 2), Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711), C127 cells (ATCC N° CRL-1804), 3T3 (ATCC N° CRL-6361), CHO (ATCC N° CCL-61), human kidney 293. (ATCC N° 45504; N° CRL-1573) and BHK (ECACC N° 84100501; N° 84111301).

c) Other mammalian host cells.

The *FLAP* gene expression in mammalian, and typically human, cells may be rendered defective, or alternatively it may be proceeded with the insertion of a *FLAP* genomic or cDNA sequence with the replacement of the *FLAP* gene counterpart in the genome of an animal cell by a *FLAP* polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of cell hosts that may be used are mammal zygotes, such as murine zygotes. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml –for BAC inserts- 3 ng/µl –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7. 4, 250 µM EDTA containing 100 mM NaCl, 30 µM spermine, and70 µM spermidine. When the DNA to

be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl et al (1993b).

Anyone of the polynucleotides of the invention, including the DNA constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC n° CRL-1821), ES-D3 (ATCC n° CRL1934 and n° CRL-11632), YS001 (ATCC n° CRL-11776), 36. 5 (ATCC n° CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells consist of primary embryonic fibroblasts that are established from tissue of day 13 to day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo et al.(1993) and are inhibited in growth by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

Transgenic Animals

The terms "transgenic animals" or "host animals" are used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a *FLAP* gene disrupted by homologous recombination with a knock out vector.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a *FLAP* coding sequence, a *FLAP* regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Generally, a transgenic animal according the present invention comprises any one of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention.

A further transgenic animals according to the invention contains in their somatic cells and/or in their germ line cells a polynucleotide comprising a biallelic marker selected from the group consisting of A1 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. Optionally said biallelic marker is selected from the group consisting of A1 to A13, A15, A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native FLAP protein, or alternatively a mutant FLAP protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the *FLAP* gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals, including knock out animals, of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to US Patents Nos. 4,873,191, issued Oct. 10, 1989, 5,464,764 issued Nov. 7, 1995; 5,789,215, issued Aug. 4, 1998; Capecchi, M.R. (1989a); Capecchi, M.R. (1989b); and Tsuzuki, T. and Rancourt, D.E. (1998), these documents being hereby incorporated by reference.

The present invention encompasses knock out vectors comprising the novel polynucleotides of the invention, as well as mammalian host cells and non-human host mammals comprising a *FLAP* gene disrupted by homologous recombination with such a knock out vector

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a *FLAP* coding sequence, a *FLAP* regulatory polynucleotide or a DNA sequence encoding a *FLAP* antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas et al.(1987). The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al.(1988).

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987). The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood et al.(1993) or by Nagy et al.(1993), the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention.

A further object of the invention consists of recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or a *FLAP* gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay et al.(1991).

XVII. Screening Of Agents Acting On The Leukotriene Pathway

In a further embodiment, the present invention also concerns a method for the screening of new agents, or candidate substances, acting on the leukotriene pathway and which may be suitable for the treatment of a patient whose DNA comprises an allele of the *FLAP* gene associated with a disease involving the leukotriene pathway, more particularly asthma.

In a preferred embodiment, the invention relates to a method for the screening of candidate substances for their ability to alter leukotriene biosynthesis, preferably to identify active candidate substances without undesired side-effects such as increased liver transaminase levels. The method comprises the following steps: a) providing a cell line, an organ, or a mammal expressing 5-LO and either a *FLAP* gene comprising alleles for one or more *FLAP*-related biallelic markers, preferably associated with a modified leukotriene pathway, more preferably with a disease involving the leukotriene pathway such as asthma, or a mutated *FLAP* gene comprising the trait cause mutation determined using the above-noted method; b) obtaining a candidate substance; and, c) testing the ability of the candidate substance to modify leukotriene biosynthesis, and particularly to interact with the 5-LO and/or with the FLAP produced by the cell line or the transgenic mammal and/or to modify the interaction between 5-LO and FLAP and/or to modulate the expression levels of FLAP.

In one embodiment of the above method, the method comprises providing a cell line, an organ, or a mammal expressing 5-LO, a FLAP gene comprising alleles for one or more *FLAP*-related biallelic markers, preferably associated with a modified leukotriene pathway, more preferably with a disease involving the leukotriene pathway such as asthma, and a mutated *FLAP* gene comprising the trait cause mutation determined using the above-noted method. Said biallelic markers may be selected from the group consisting of A1 to A28, and the complements thereof; Optionally, said *FLAP*-related biallelic marker may be selected from the group consisting of A1 to A13, A15, and A17 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A1 to A10 and A22 to A28, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said biallelic markers are selected from the group consisting of A14 or A16, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. In a preferred embodiment, said biallelic markers are selected from the group consisting of A2, A14, A16, A18, A19, A22, and A23, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. In another preferred embodiment said biallelic markers are selected from the group consisting of A14 and A19, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith. In more preferred embodiment, said biallelic markers comprise the biallelic marker A19, and the complement thereof, or optionally the biallelic markers in linkage disequilibrium therewith.

A candidate substance is a substance which can interact with or modulate, by binding or other intramolecular interactions, 5-LO or FLAP. Such substances may be potentially interesting for patients who are not responsive to existing drugs. Screening may be effected using either in vitro methods or in vivo methods.

In vitro methods can be carried out in numerous ways such as on transformed cells which express the considered alleles of the *FLAP* gene through 5-lipoxygenase activation and leukotriene synthesis measurement or on FLAP encoded by the considered allelic variant of *FLAP* through FLAP binding assays.

Screening assays of the present invention generally involve determining the ability of a candidate substance to affect the activity of 5-LO or FLAP, such as the screening of candidate substances to identify those that inhibit or otherwise modify the function of 5-LO or FLAP in the leukotriene pathway.

One method of drug screening utilizes eukaryotic host cells which are stably transformed with recombinant polynucleotides expressing 5-LO and the considered alleles of the FLAP gene. Such cells, either in viable or fixed form, can be used for standard binding assays. One can measure, for example, the formation of products of the leukotriene pathway such as LTB_4 synthesis or

examine the degree to which the formation of such products is interfered with by the agent being tested.

Typically, this method includes preparing transformed cells which express 5-LO and different forms of FLAP encoded by DNA sequences containing particular alleles of one or more of the biallelic markers and/or mutations described above. This is followed by testing the cells expressing the 5-LO and FLAP with a candidate substance to determine the ability of the substance to affect the leukotriene pathway function, in order to identify those which affect the enzymatic activity of 5-LO or the activity of FLAP, and which thus can be suitable for use in humans.

Typical examples of such drug screening assays are provided below. It is to be understood that the parameters set forth in these examples can be modified by the skilled person without undue experimentation.

Screening for 5-LO inhibitors

Drug effects can be evaluated by assessing the 5-LO products generated by cells expressing both the 5-LO gene and the considered allele of the *FLAP* gene. Eukaryotic cells previously transformed with appropriate vectors as described previously and expressing 5-LO and the allele of the *FLAP* gene under study are harvested by centrifugation (300 g, 5 min, and room temperature) and washed with an appropriate buffer. The cells are then resuspended in buffer, pre-warmed at 37°C, preferably at a cell density of 5×10^6 cells/ml. Aliquots of the cell suspension are incubated with the considered drug for preferably 5 min at 37°C. Reaction is initiated by the addition of calcium ionophore A23187 and arachidonic acid. Following incubation at 37°C, reaction is stopped by adding methanol containing prostaglandin B2 as an internal standard for HPLC analysis. 5-LO reaction products are extracted into chloroform, dried under a stream of nitrogen, and resuspended in HPLC solvent. The samples are analyzed by reverse-phase HPLC using preferably an isocratic solvent system of methanol/water/acetic acid (75:25:0.01). The elution is monitored at preferably 270 and 234 nm. 5-LO products are quantitated by comparison of peak areas to those of standard curves of authentic standards, and corrected for minor differences in extraction efficiency determined using the prostaglandin B2 internal standard. This method is described in more detail in Dixon et al. (1990) and Abramovitz et al. (1993), the disclosures of which are incorporated herein by reference.

Screening for FLAP inhibitors

The FLAP protein or portions thereof described above may be used in drug screening procedures to identify molecules which are agonists, antagonists, or inhibitors of FLAP activity. The FLAP protein or portion thereof used in such analyses may be free in solution or linked to a solid support. Alternatively, FLAP protein or portions thereof can be expressed on a cell surface. The cell may naturally express the FLAP protein or portion thereof or, Alternatively, the cell may express the FLAP protein or portion thereof from an expression vector such as those described above.

In one method of drug screening, eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides in order to express the FLAP protein or a portion thereof are used in conventional competitive binding assays or standard direct binding assays. For example, the formation of a complex between the FLAP protein or a portion thereof and the agent being tested may be measured in direct binding assays. Alternatively, the ability of a test agent to prevent formation of a complex between the FLAP protein or a portion thereof and a known ligand may be measured.

For example, a FLAP inhibitor binding assay can be based on the observation that MK-886, an indole leukotriene biosynthesis inhibitor, binds with high affinity and specificity to FLAP. Binding of the considered drug to FLAP can be assessed by a competition experiments with a radiolabeled analog of MK-886, ^{125}I-L-691-831. A suspension of cells expressing the considered allele of FLAP containing preferably 2 10^7 cells is centrifuged at 500 x g for 10 min. The pelleted cells are then resuspended in lysis buffer. This suspension is sonicated on ice by three 20 sec bursts. Cell lysis is checked visually. Binding is initiated by addition of cell lysis samples to wells containing ^{125}I-L-691-831 and either the considered drug or nothing (control). The plate is incubated for 20 min at room temperature. The samples are then filtered and washed. Bound ^{125}I-L-691-831 is determined in a counter. Specific drug binding is defined as the difference between binding in the absence and the presence of the considered drug. This FLAP binding assay is described with more details in Charleson et al. (1992).

Alternatively, the high throughput screening techniques disclosed in published PCT application WO 84/03564 may be used. In such techniques, large numbers of small peptides to be tested for FLAP binding activity are synthesized on a surface and affixed thereto. The test peptides are contacted with the FLAP protein or a portion thereof, followed by a wash step. The amount of FLAP protein or portion thereof which binds to the test compound is quantitated using conventional techniques.

In some methods, FLAP protein or a portion thereof may be fixed to a surface and contacted with a test compound. After a washing step, the amount of test compound which binds to the FLAP protein or portion thereof is measured.

Screening for inhibitors of the interaction between 5-LO and FLAP

Drug effects can be evaluated through the assessment of the interaction between 5-LO and FLAP proteins.

Interaction between 5-LO and FLAP protein may be assessed using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No K1604-1, Clontech). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No K1604-1, Clontech) nucleic acids encoding the FLAP protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. 5-LO cDNA or a portion thereof is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on

each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between FLAP and 5-LO proteins.

In another method, affinity columns containing the FLAP protein or a portion thereof can be constructed. In some versions of this method the affinity column contains chimeric proteins in which the FLAP protein or a portion thereof is fused to glutathione S-transferase. 5-LO protein is applied to the affinity column. The 5-LO protein retained on the affinity column can be measured and can allow assessment of the interaction between FLAP and 5-LO proteins.

Association between 5-LO and FLAP proteins can also be assessed by using an Optical Biosensor as described in Edwards et Leatherbarrow, (1997). The main advantage of the method is that it allows the determination of the association rate. Typically a FLAP molecule is linked to the sensor surface (through a carboxymethl dextran matrix) and a sample of 5-LO molecules is placed in contact with the FLAP molecules. The binding of a 5-LO molecule to the FLAP molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). Hence, the effect of candidate drug on the association between FLAP and 5-LO proteins can be easily measured.

Screening for expression modifiers

The screening of expression modifiers is important as it can be used for detecting modifiers specific to one allele or a group of alleles of the *FLAP* gene. The alteration of *FLAP* expression in response to a modifier can be determined by administering or combining the candidate modifier with an expression system such as animal, or cell, and in *in vitro* transcription assay.

The effect of the modifier on *FLAP* transcription and /or steady state mRNA levels can also be determined. As with the basic expression levels, tissue specific interactions are of interest. Correlations are made between the ability of an expression modifier to affect FLAP activity, and the presence of the targeted polymorphisms. A panel of different modifiers may be screened in order to determine the effect under a number of different conditions.

Expression levels and patterns of *FLAP* may be analyzed by solution hybridization with long probes as described in International Patent Application No WO 97/05277, the entire contents of which are incorporated herein by reference. Briefly, the *FLAP* cDNA or the *FLAP* genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the *FLAP* insert comprises at least 100 or more consecutive nucleotides of the genomic DNA sequence or the cDNA sequences, particularly those comprising at least one of the biallelic markers of the present invention or those encoding mutated FLAP. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this

doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

Quantitative analysis of *FLAP* gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of a plurality of nucleic acids of sufficient length to permit specific detection of expression of mRNAs capable of hybridizing thereto. For example, the arrays may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The arrays may include the *FLAP* genomic DNA, the *FLAP* cDNA sequences or the sequences complementary thereto or fragments thereof, particularly those comprising at least one of the biallelic markers of the present invention or those encoding mutated FLAP. Preferably, the fragments are at least 15 nucleotides in length. In other embodiments, the fragments are at least 25 nucleotides in length. In some embodiments, the fragments are at least 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. In another preferred embodiment, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

For example, quantitative analysis of *FLAP* gene expression may be performed with a complementary DNA microarray as described by Schena et al. (1995 and 1996). Full length *FLAP* cDNAs or fragments thereof are amplified by PCR and arrayed from a 96-well microtiter plate onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of *FLAP* gene expression may also be performed with full length *FLAP* cDNAs or fragments thereof in complementary DNA arrays as described by Pietu et al. (1996). The full length *FLAP* cDNA or fragments thereof is PCR amplified and spotted on membranes. Then,

mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis using the *FLAP* genomic DNA, the *FLAP* cDNA, or fragments thereof can be done through high density nucleotide arrays as described by Lockhart et al. (1996) and Sosnowsky et al. (1997). Oligonucleotides of 15-50 nucleotides from the sequences of the *FLAP* genomic DNA, the *FLAP* cDNA sequences, particularly those comprising at least one of the biallelic markers of the present invention or those encoding mutated FLAP, or the sequences complementary thereto, are synthesized directly on the chip (Lockhart et al., 1996) or synthesized and then addressed to the chip (Sosnowski et al., 1997). Preferably, the oligonucleotides are about 20 nucleotides in length.

FLAP cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart et al., supra and application of different electric fields (Sosnowsky et al., 1997)., the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of *FLAP* mRNA.

Screening using transgenic animals

In vivo methods can utilize transgenic animals for drug screening. Nucleic acids including at least one of the biallelic polymorphisms of interest can be used to generate genetically modified non-human animals or to generate site specific gene modifications in cell lines. The term "transgenic" is intended to encompass genetically modified animals having a deletion or other knock-out of *FLAP* gene activity, having an exogenous *FLAP* gene that is stably transmitted in the host cells, or having an exogenous *FLAP* promoter operably linked to a reporter gene. Transgenic animals may be made through homologous recombination, where the *FLAP* locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include for example plasmids, retroviruses and other animal viruses, and YACs. Of interest are transgenic mammals e.g. cows, pigs, goats, horses, and particularly rodents such as rats and mice. Transgenic animals allow to study both efficacy and toxicity of the candidate drug.

XVIII. Computer-Related Embodiments

As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any one of the following: a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1 of the

following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069; b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a C at position 16348of SEQ ID No 1; c) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329; d) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1; e) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2, wherein said contiguous span comprises a T at position 197, an A at position 453, or a G at position 779 of SEQ ID No 2; and f) a nucleotide sequence complementary to any one of the preceding nucleotide sequences.

The "nucleic acid codes of the invention" further encompass nucleotide sequences homologous to a contiguous span of at least 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of the following nucleotide position range: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069 of SEQ ID No 1, and sequences complementary to all of the preceding sequences. Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention. It will be appreciated that the nucleic acid codes of the invention can be represented in the traditional single character format (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman & Co., New York.) or in any other format or code which records the identity of the nucleotides in a sequence.

As used herein the term "polypeptide codes of the invention" encompass the polypeptide sequences comprising a contiguous span of at least 6, 8, 10, 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 of SEQ ID No 3. It will be appreciated that the polypeptide codes of the invention can be represented in the traditional single character format or three letter format (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman & Co., New York.) or in any other format or code which records the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a

process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, DVD, RAM, or ROM as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which contain the sequence information described herein. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to store and/or analyze the nucleotide sequences of the nucleic acid codes of the invention, the amino acid sequences of the polypeptide codes of the invention, or other sequences. The computer system preferably includes the computer readable media described above, and a processor for accessing and manipulating the sequence data.

Preferably, the computer is a general purpose system that comprises a central processing unit (CPU), one or more data storage components for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system includes a processor connected to a bus which is connected to a main memory, preferably implemented as RAM, and one or more data storage devices, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system further includes one or more data retrieving devices for reading the data stored on the data storage components. The data retrieving device may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, a hard disk drive, a CD-ROM drive, a DVD drive, etc. In some embodiments, the data storage component is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device. Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention, or the amino acid sequences of the polypeptide codes of the invention (such as search tools, compare tools, modeling tools, etc.) may reside in main memory during execution.

In some embodiments, the computer system may further comprise a sequence comparer for comparing the nucleic acid codes of the invention or polypeptide codes of the invention stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals the sequences or structures of which are stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of the nucleic acid codes of the invention, or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of the invention, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of the invention.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of the invention through the use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences .

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide

sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention.

An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention.

The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the invention. The following list is

intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag et al., 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius^2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Throughout this application, various publications, patents, and published patent applications are cited. The disclosures of the publications, patents, and published patent specifications referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

EXAMPLES

Example 1

Detection of *FLAP* biallelic markers: DNA extraction

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M;
- 200 µl SDS 10%; and
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

Example 2

Detection of the biallelic markers: amplification of genomic DNA by PCR

The amplification of specific genomic sequences of the DNA samples of example 1 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

Final volume	25 µl
DNA	2 ng/µl
$MgCl_2$	2 mM
dNTP (each)	200 µM
primer (each)	2.9 ng/µl
Ampli Taq Gold DNA polymerase	0.05 unit/µl
PCR buffer (10x = 0.1 M Tris-HCl pH8.3 0.5M KCl)	1x

Each pair of first primers was designed using the sequence information of the *FLAP* gene (GenBank 182657, Kennedy et al. 1991 incorporated herein by reference) and the OSP software (Hillier & Green, 1991). These first primers had about 20 nucleotides in length and their respective sequences are disclosed in Table 1.

Table 1

Amplicon	Position range of the amplicon in SEQ ID No 1		PU	Position range of amplification primer in SEQ ID No 1		RP	Complementary position range of amplification primer in SEQ ID No 1	
10-517	3851	4189	B15	3851	3869	C15	4171	4189
10-518	4120	4390	B16	4120	4138	C16	4372	4390
10-253	4373	4792	B1	4373	4391	C1	4773	4792
10-499	4814	5043	B2	4814	4833	C2	5026	5043
10-500	4956	5422	B3	4956	4972	C3	5405	5422
10-522	5524	5996	B17	5524	5542	C17	5978	5996
10-503	6218	6672	B4	6218	6235	C4	6652	6672
10-504	6522	6790	B5	6522	6539	C5	6772	6790
10-204	7120	7574	B6	7120	7137	C6	7557	7574
10-32	7513	7933	B7	7513	7531	C7	7914	7933
10-33	16114	16533	B8	16114	16132	C8	16515	16533
10-34	24072	24425	B9	24072	24089	C9	24408	24425
10-35	27978	28401	B10	27978	27995	C10	28384	28401
10-36	36020	36465	B11	36020	36039	C11	36446	36465
10-498	36318	36669	B12	36318	36337	C12	36652	36669
12-629	38441	38840	B13	38441	38460	C13	38820	38840
12-628	42233	42749	B14	42233	42253	C14	42731	42749

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence: TGTAAAACGACGGCCAGT (SEQ ID No 14); primers RP contain the following RP 5' sequence: CAGGAAACAGCTATGACC (SEQ ID No 15).

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 94°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 94°C, 55°C for 1 min, and 30 sec at 72°C. For final elongation, 7 min at 72°C end the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

Example 3

Detection of the biallelic markers:

Sequencing of amplified genomic DNA and identification of polymorphisms

The sequencing of the amplified DNA obtained in example 2 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined.

The sequence data were further evaluated for polymorphisms by detecting the presence of biallelic markers among the pooled amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position.

17 fragments of amplification were analyzed. In these segments, 28 biallelic markers were detected. The localization of the biallelic markers was as shown in Table 2.

Table 2

Amplicon	BM	Marker Name	Freq. of all2	Localization in *FLAP* gene	Polymorphism		BM position in SEQ ID		Position of 47mers in		47 mers name
					all1	all2	No 1	No 2	SEQ ID No 1		
10-517	A25	10-517-100		5'regulatory	G	C	3950		3927	3973	P25
10-518	A26	10-518-125		5'regulatory	G	T	4243		4220	4266	P26
10-518	A27	10-518-194		5'regulatory	A	G	4312		4289	4335	P27
10-253	A1	10-253-118		5'regulatory	A	G	4490		4467	4513	P1
10-253	A2	10-253-298	4.57	5'regulatory	G	C	4670		4647	4693	P2
10-253	A3	10-253-315		5'regulatory	C	T	4687		4664	4710	P3
10-499	A4	10-499-155		5'regulatory	A	G	4968		4945	4991	P4
10-500	A5	10-500-185		5'regulatory	C	T	5140		5117	5163	P5
10-500	A6	10-500-258		5'regulatory	G	T	5213		5190	5236	P6
10-500	A7	10-500-410		5'regulatory	A	G	5364		5341	5387	P7
10-522	A28	10-522-71		5'regulatory	A	G	5594		5571	5617	P28
10-503	A8	10-503-159		5'regulatory	G	T	6370		6347	6393	P8
10-504	A9	10-504-172		5'regulatory	A	T	6693		6670	6716	P9
10-504	A10	10-504-243		5'regulatory	A	C	6763		6740	6786	P10
10-204	A11	10-204-326	6.63	5'regulatory	A	G	7445		7422	7468	P11
10-32	A12	10-32-357	33.45	Intron 1	A	C	7870		7847	7893	P12
10-33	A13	10-33-175	2.3	Exon 2	C	T	16288	197	16265	16311	P13
10-33	A14	10-33-234	43.98	Intron 2	A	C	16347		16324	16370	P14
10-33	A15	10-33-270		Intron 2	A	G	16383		16360	16406	P15
10-33	A16	10-33-327	24.26	Intron 2	C	T	16440		16417	16463	P16
10-34	A17	10-34-290		Intron 3	G	T	24361		24338	24384	P17
10-35	A18	10-35-358	31.25	Intron 4	G	C	28336		28313	28359	P18
10-35	A19	10-35-390	22.98	Intron 4	C	T	28368		28345	28391	P19
10-36	A20	10-36-164		Exon 5 V127→I	A	G	36183	453	36160	36206	P20
10-498	A21	10-498-192		Exon 5	A	G	36509	779	36486	36532	P21
12-629	A22	12-629-241	28.3	3'regulatory	G	C	38681		38658	38704	P22
12-628	A24	12-628-311		3'regulatory	T	C	42440		42417	42463	P24
12-628	A23	12-628-306	10.27	3'regulatory	G	A	42445		42422	42468	P23

BM refers to "biallelic marker". All1 and all2 refer respectively to allele 1 and allele 2 of the biallelic marker. "Freq. Of all2" refers to the frequency of the allele 2 in percentage in Caucasian US control population, except for the biallelic marker 10-204/326 for which the population is the French Caucasian controls. Frequencies corresponded to a population of random blood donors from French Caucasian origin.

The polymorphisms A14 (10-33-234) and A16 (10-33-327) have been observed in Kennedy et al, 1991. However, their frequencies in the population was unknown, therefore they can not be considered validated biallelic markers, until the results of the present inventors were obtained.

Example 4

Validation of the polymorphisms through microsequencing

The biallelic markers identified in example 3 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each individual DNA sample described in Example 1.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing had about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base. Their sequences are disclosed in Table 3 below.

Table 3

Marker Name	Biallelic Marker	Mis. 1	Position range of microsequencing primer mis 1 in SEQ ID No 1		Mis. 2	Complementary position range of microsequencing primer mis. 2 in SEQ ID No 1	
10-517-100	A25	D25	3930	3949	E25	3951	3970
10-518-125	A26	D26	4223	4242	E26	4244	4263
10-518-194	A27	D27	4292	4311	E27	4313	4332
10-253-118	A1	D1	4470	4489	E1	4491	4510
10-253-298	A2	D2	4650	4669	E2	4671	4690
10-253-315	A3	D3	4667	4686	E3	4688	4707
10-499-155	A4	D4	4948	4967	E4	4969	4988
10-500-185	A5	D5	5120	5139	E5	5141	5160
10-500-258	A6	D6	5193	5212	E6	5214	5233
10-500-410	A7	D7	5344	5363	E7	5365	5384
10-522-71	A28	D28	5574	5593	E28	5595	5614
10-503-159	A8	D8	6350	6369	E8	6371	6390
10-504-172	A9	D9	6673	6692	E9	6694	6713
10-504-243	A10	D10	6743	6762	E10	6764	6783
10-204-326	A11	D11	7425	7444	E11	7446	7465
10-32-357	A12	D12	7850	7869	E12	7871	7890
10-33-175	A13	D13	16268	16287	E13	16289	16308
10-33-234	A14	D14	16327	16346	E14	16348	16367
10-33-270	A15	D15	16363	16382	E15	16384	16403
10-33-327	A16	D16	16420	16439	E16	16441	16460
10-34-290	A17	D17	24341	24360	E17	24362	24381
10-35-358	A18	D18	28316	28335	E18	28337	28356
10-35-390	A19	D19	28348	28367	E19	28369	28388
10-36-164	A20	D20	36163	36182	E20	36184	36203
10-498-192	A21	D21	36489	36508	E21	36510	36529
12-629-241	A22	D22	38661	38680	E22	38682	38701
12-628-311	A24	D24	42420	42439	E24	42441	42460
12-628-306	A23	D23	42425	42444	E23	42446	42465

Mis 1 and Mis 2 respectively refer to microsequencing primers which hybridized with the non-coding strand of the *FLAP* gene or with the coding strand of the *FLAP* gene.

The microsequencing reaction was performed as follows:

5 µl of PCR products were added to 5 µl purification mix 2U SAP (Shrimp alkaline phosphate) (Amersham E70092X)); 2U Exonuclease I (Amersham E70073Z); 1 µl SAP buffer (200 mM Tris-HCl pH8, 100 mM $MgCl_2$) in a microtiter plate. The reaction mixture was incubated 30 minutes at 37°C, and denatured 10 minutes at 94°C afterwards. To each well was then added 20 µl of microsequencing reaction mixture containing: 10 pmol microsequencing oligonucleotide (19mers, GENSET, crude synthesis, 5 OD), 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs complementary to the nucleotides at the polymorphic site corresponding to both polymorphic bases (11.25 nM TAMRA-ddTTP; 16.25 nM ROX-ddCTP; 1.675 nM REG-ddATP; 1.25 nM RHO-ddGTP; Perkin Elmer, Dye Terminator Set 401095). After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The microtiter plate was centrifuged 10 sec at 1500 rpm. The unincorporated dye terminators were removed by precipitation with 19 µl $MgCl_2$ 2mM and 55 µl 100 % ethanol. After 15 minute incubation at room temperature, the microtiter plate was centrifuged at 3300 rpm 15 minutes at 4°C. After discarding the supernatants, the microplate was evaporated to dryness under reduced pressure (Speed Vac); samples were resuspended in 2.5 µl formamide EDTA loading buffer and heated for 2 min at 95°C. 0.8 µl microsequencing reaction were loaded on a 10 % (19:1) polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Example 5

Association study between asthma and the biallelic markers of the *FLAP* gene: collection of DNA samples from affected and non-affected individuals

The disease trait followed in this association study was asthma, a disease involving the leukotriene pathway.

The asthmatic population corresponded to 297 individuals that took part in a clinical study for the evaluation of the anti-asthmatic drug Zileuton. More than 90 % of these 297 asthmatic individuals had a Caucasian ethnic background.

The control population corresponded to unaffected individuals. In this association study, either Caucasian French population (190 individuals) or Caucasian US population (286 individuals) is used as control population. The preferred control population is the Caucasian US population since the asthmatic population essentially comprises US individuals.

Example 6

Association study between asthma and the biallelic markers of the *FLAP* gene: Genotyping of affected and control individuals

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of the above described biallelic markers in each of these populations.

Allelic frequencies of the above-described biallelic markers in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in examples 2 and 4 using the described PCR and microsequencing primers.

Example 7

Association study between asthma and the biallelic markers of the *FLAP* gene

A) Association studies for asthma gene with Caucasian French control population

This association study uses 293 asthmatic individuals and 185 Caucasian French controls.

As shown in Figure 2 (A), markers 10-32/357 and 10-35/390 presented a strong association with asthma, this association being highly significant (pvalue = $1.95x10^{-3}$ for marker 10-32/357 and $1.75x10^{-3}$ for marker 10-35-390). The two markers 10-32/357 and 10-35/390 can be then used in diagnostics with a test based on each marker. Two other markers showed moderate association when tested independently, namely 33/234, and 35/358.

B) Association studies for asthma gene with Caucasian US control population

This association study uses 297 asthmatic individuals and 286 Caucasian US controls.

As shown in Figure 2 (B), the biallelic marker 10-35/390 presented a strong association with asthma, this association being highly significant (pvalue = $2.29x10^{-3}$). The two markers 10-32/357 and 10-33/234 showed weak association when tested independently.

The biallelic marker 10-35/390 is located in the genomic sequence of *FLAP*. Therefore, the association studies results show that a polymorphism of the *FLAP* gene seems to be related to asthma. The biallelic marker 10-35/390 can be then used in diagnostics with a test based on this marker or on a combination of biallelic markers comprising this marker.

Example 8

Association studies: Haplotype frequency analysis

One way of increasing the statistical power of individual markers, is by performing haplotype association analysis.

Haplotype analysis for association of *FLAP* markers and asthma was performed by estimating the frequencies of all possible haplotypes comprising biallelic markers selected from the

group consisting of 10-253/298, 10-32/357, 10-33/175, 10-33/234, 10-33/327, 10-35/358, 10-35/390, 12-628/306, and 12-629/241 in the asthmatic and Caucasian US control populations described in Example 7, and comparing these frequencies by means of a chi square statistical test (one degree of freedom). Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier L & Slatkin M, 1995), using the EM-HAPLO program (Hawley ME, Pakstis AJ & Kidd KK, 1994).

The most significant haplotypes obtained are shown in Figure 3.

The preferred two-markers haplotypes, described in figure 3 as HAP1 to HAP 7, comprise either the marker 10-33/234 (allele A) or the marker 10-35/390 (allele T). The more preferred two-markers haplotype HAP1 (A at 10-33/234 and T at 10-35/390) presented a p-value of $8.2x10^{-4}$ and an odd-ratio of 1.61. Estimated haplotype frequencies were 28.3% in the cases and 19.7 % in the US controls. Two other two-haplotypes HAP2 (A at 10-33/234 and G at 12-629/241) and HAP3 (T at 10-33/327 and T at 10-33/390) presented respectively a p-value of $1.6x10^{-3}$ and $1.8x10^{-3}$, an odd-ratio of 1.65 and 1.53 and haplotypes frequencies of 0.305 and 0.307 for asthmatic population and of 0.210 and 0.224 for US control population.

Preferred three-markers haplotypes comprise the marker 10-33/234 (allele A) and the marker 10-35/390 (allele T): HAP37, HAP38, HAP39 and HAP41. The more preferred three-markers haplotype HAP37 (A at 10-33/234, T at 10-33/390 and C at 12-628/306) presented a p-value of $8.6x10^{-4}$ and an odd-ratio of 1.76. Estimated haplotype frequencies were 26.5 % in the cases and 17.1 % in the US controls. A further three-markers haplotype HAP40 (A at 10-33/234, C at 12-628/306 and G at 12-629/241) is also significant.

Four-markers haplotypes (HAP121 to HAP125), five-markers haplotypes (HAP 247 and 248) and a six-markers haplotype (HAP373) showed significant p-values. They all comprise the marker 10-33/234 (allele A) and the marker 10-35/390 (allele T), except the haplotype HAP124 which does not comprise the marker 10-35/390. The other markers are chosen from the group consisting of 10-235/298 (allele C), 10-35/358 (allele G), 12-628/306 (allele C) and 12-629/241 (allele G).

The more preferred haplotype comprising A at 10-33/234 and T at 10-35/390 (HAP1 in figure 3) is also significant in a haplotype frequency analysis with asthmatic population and Caucasian French controls. Indeed, this haplotype presented a p-value of $2.7x10^{-3}$ and an odd-ratio of 1.67. Estimated haplotype frequencies were 28.3 % in the cases and 19.2 % in the French controls (see Figure 4).

The haplotype HAP1 is the more preferred haplotype of the invention. It can be used in diagnosis of asthma. Moreover, most of the significant haplotypes associated with asthma comprise the biallelic marker 10-35/390 (allele A) and could also be used in diagnosis.

The statistical significance of the results obtained for the haplotype analysis was evaluated by a phenotypic permutation test reiterated 1000 or 10,000 times on a computer. For this computer

simulation, data from the asthmatic and control individuals were pooled and randomly allocated to two groups which contained the same number of individuals as the case-control populations used to produce the data summarized in figure 3. A haplotype analysis was then run on these artificial groups for the 2 markers included in the haplotype HAP1 which, showed the strongest association with asthma. This experiment was reiterated 1000 and 10,000 times and the results are shown in figure 4. These results demonstrate that among 1000 iterations none and among 10,000 iterations only 1 of the obtained haplotypes had a p-value comparable to the one obtained for the haplotype HAP1. These results clearly validate the statistical significance of the association between this haplotype and asthma.

Example 9

Preparation of Antibody Compositions to the 127-Ile Variant of FLAP

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the FLAP protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the FLAP protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., (1975) or derivative methods thereof. Also see Harlow, E., and D. Lane. 1988.

Briefly, a mouse is repetitively inoculated with a few micrograms of the FLAP protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the FLAP protein or a portion thereof can be prepared by immunizing suitable non-human animal with the FLAP protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for FLAP concentration can be

used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein by the one skilled in the art without departing from the spirit and scope of the invention.

REFERENCES

Abbondanzo SJ et al., 1993, Methods in Enzymology, Academic Press, New York, pp. 803-823 / Abramovitz M, Wong E, Cox ME, Richardson CD, Li C, & Vickers PJ, *Eur J Biochem* 1993;215(1):105-111 / Altschul et al., 1990, J. Mol. Biol. 215(3):403-410 / Altschul et al., 1993, Nature Genetics 3:266-272 / Altschul et al., 1997, Nuc. Acids Res. 25:3389-3402 / Anton M. et al., 1995, J. Virol., 69 : 4600-4606 / Araki K et al. (1995) *Proc. Natl. Acad. Sci. USA.* 92(1):160-4. / Ausubel et al. (1989)Current Protocols in Molecular Biology, Green Publishing

Associates and Wiley Interscience, N.Y. / Baubonis W. (1993) *Nucleic Acids Res.* 21(9):2025-9. / Beaucage et al., *Tetrahedron Lett* 1981, 22: 1859-1862 / Bradley A., 1987, Production and analysis of chimaeric mice. *In*: E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113. / Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;68:109-151 / Brutlag et al. Comp. App. Biosci. 6:237-245, 1990 / Capecchi, M.R. (1989a) *Science*, 244:1288-1292 / Capecchi, M.R. (1989b) *Trends Genet.*, 5:70-76 / Chai H. et al. (1993) *Biotechnol. Appl. Biochem.*18:259-273. / Charleson S, Prasit P, Leger S, Gillard JW, Vickers PJ, Mancini JA, Charleson P, Guay J, Ford-Hutchinson AW, Evans JF, *Mol Pharmacol* 1992;41(5):873-879 / Chee et al. (1996) *Science.* 274:610-614. / Chen et al. (1987) *Mol. Cell. Biol.* 7:2745-2752. / Chou J.Y., 1989, Mol. Endocrinol., 3: 1511-1514. / Clark A.G. (1990) *Mol. Biol. Evol.* 7:111-122. / Compton J. (1991) *Nature.* 350(6313):91-92. / Davis L.G., M.D. Dibner, and J.F. Battey, Basic Methods in Molecular Biology, ed., Elsevier Press, NY, 1986 / Dempster et al., (1977) *J. R. Stat. Soc.*, 39B:1-38. / Dixon et al., 1988, *PNAS* 85, pp. 416-420 / Dixon RA, Diehl RE, Opas E, Rands E, Vickers PJ, Evans JF, Gillard JW, & Miller DK, *Nature* 1990;343(6255):282-284 / Eckner R. et al. (1991) *EMBO J.* 10:3513-3522. / Edwards et Leatherbarrow, *Analytical Biochemistry*, 246, 1-6 (1997) / Engvall, E., Meth. Enzymol. 70:419 (1980) / Excoffier L. and Slatkin M. (1995) *Mol. Biol. Evol.*, 12(5): 921-927. / Feldman and Steg, 1996, Medecine/Sciences, synthese, 12:47-55 / Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980) / Flotte et al. (1992) *Am. J. Respir. Cell Mol. Biol.* 7:349-356. / Fodor et al. (1991) *Science* 251:767-777. / Fraley et al. (1979) *Proc. Natl. Acad. Sci. USA.* 76:3348-3352. / Fuller S. A. et al. (1996) *Immunology in Current Protocols in Molecular Biology*, Ausubel et al.Eds, John Wiley & Sons, Inc., USA. / Furth P.A. et al. (1994) *Proc. Natl. Acad. Sci USA.* 91:9302-9306. / Geysen H. Mario et al. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002 / Ghosh and Bacchawat, 1991, *Targeting of liposomes to hepatocytes*, IN: *Liver Diseases, Targeted diagnosis and therapy using specific rceptors and ligands*. Wu et al. Eds., Marcel Dekeker, New York, pp. 87-104. / Gonnet et al., 1992, Science 256:1443-1445 / Gopal (1985) *Mol. Cell. Biol.*, 5:1188-1190. / Gossen M. et al. (1992) *Proc. Natl. Acad. Sci. USA.* 89:5547-5551. / Gossen M. et al. (1995) *Science.* 268:1766-1769. / Graham et al. (1973) *Virology* 52:456-457. / Grompe, M. (1993) *Nature Genetics.* 5:111-117. / Grompe, M. et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:5855-5892. / Gu H. et al. (1993) *Cell* 73:1155-1164. / Gu H. et al. (1994) *Science* 265:103-106. / Guatelli J C et al. *Proc. Natl. Acad. Sci. USA.* 35:273-286. / Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS, *Nat Genet* 1996;14(4):441-447 / Haff L. A. and Smirnov I. P. (1997) *Genome Research*, 7:378-388. / Hames B.D. and Higgins S.J. (1985) *Nucleic Acid Hybridization: A Practical Approach.* Hames and Higgins Ed., IRL Press, Oxford. / Harju L, Weber T, Alexandrova L, Lukin M, Ranki M, Jalanko A, *Clin Chem* 1993;39(11Pt 1):2282-2287 / Harland et al. (1985) *J. Cell. Biol.* 101:1094-1095. / Harlow, E., and D. Lane. 1988. Antibodies A

Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242 / Hawley M.E. et al. (1994) *Am. J. Phys. Anthropol.* 18:104. / Henikoff and Henikoff, 1993, Proteins 17:49-61 / Higgins et al., 1996, Methods Enzymol. 266:383-402 / Hillier L. and Green P. *Methods Appl.*, 1991, 1: 124-8. / Hoess et al. (1986) *Nucleic Acids Res.* 14:2287-2300. / Huang L. et al. (1996) *Cancer Res* 56(5):1137-1141. / Huygen et al. (1996) *Nature Medicine.* 2(8):893-898. / Julan et al. (1992) *J. Gen. Virol.* 73:3251-3255. / Kanegae Y. et al., *Nucl. Acids Res.* 23:3816-3821. / Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268 / Kennedy BP, Diehl RE, Boie Y, Adam M, Dixon RA, *J Biol Chem* 1991;266(13):8511-8516 / Kim U-J. et al. (1996) *Genomics* 34:213-218. / Klein et al. (1987) *Nature.* 327:70-73. / Kohler, G. and Milstein, C., Nature 256:495 (1975) / Koller et al. (1992) *Annu. Rev. Immunol.* 10:705-730. / Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M, Gingeras TR, *Nat Med* 1996;2(7):753-759 / Landegren U. et al. (1998) *Genome Research*, 8:769-776. / Lange K. (1997) *Mathematical and Statistical Methods for Genetic Analysis.* Springer, New York. / Lenhard T. et al. (1996) *Gene.* 169:187-190. / Linton M.F. et al. (1993) *J. Clin. Invest.* 92:3029-3037. / Liu Z. et al. (1994) *Proc. Natl. Acad. Sci. USA.* 91: 4528-4262. / Livak KJ, Hainer JW, *Hum Mutat* 1994;3(4):379-385 / Livak et al., *Nature Genetics*, 9:341-342, 1995 / Lockhart et al. *Nature Biotechnology* 14: 1675-1680, 1996 / Mansour S.L. et al. (1988) *Nature.* 336:348-352. / Manz et al., *Adv in Chromatogr* 1993; 33:1-66 / Marshall R. L. et al. (1994) *PCR Methods and Applications.* 4:80-84. / McCormick et al. (1994) *Genet. Anal. Tech. Appl.* 11:158-164. / McLaughlin B.A. et al. (1996) *Am. J. Hum. Genet.* 59:561-569. / Muzyczka et al. (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129. / Nada S. et al. (1993) *Cell* 73:1125-1135. / Nagy A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 8424-8428. / Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;68:90-98 / Neda et al. (1991) *J. Biol. Chem.* 266:14143-14146. / Newton et al. (1989) *Nucleic Acids Res.* 17:2503-2516. / Nickerson D.A. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927. / Nicolau C. et al., 1987, Methods Enzymol., 149:157-76. / Nicolau et al. (1982) *Biochim. Biophys. Acta.* 721:185-190. / Nyren P, Pettersson B, Uhlen M, *Anal Biochem* 1993;208(1):171-175 / O'Reilly et al. (1992) *Baculovirus Expression Vectors: A Laboratory Manual.* W. H. Freeman and Co., New York. / Ohno et al. (1994) *Science.* 265:781-784. / Orita et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.*86: 2776-2770. / Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973) / Pastinen et al., *Genome Research* 1997; 7:606-614 / Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85(8):2444-2448 / Pease S. ans William R.S., 1990, Exp. Cell. Res., 190: 209-211. / Perlin et al. (1994) *Am. J. Hum. Genet.* 55:777-787. / Pietu et al. *Genome Research* 6:492-503, 1996 / Potter et al. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81(22):7161-7165. / Reid L.H. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:4299-4303. / Robertson E., 1987, Embryo-derived stem cell lines. In: E.J. Robertson Ed. *Teratocarcinomas and embrionic stem cells: a practical approach.* IRL Press, Oxford, pp. 71. / Roth J.A. et al. (1996) *Nature Medicine.* 2(9):985-991. / Roux et al.

(1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:9079-9083. / Ruano et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:6296-6300. / Sambrook, J., Fritsch, E.F., and T. Maniatis. (1989) *Molecular Cloning: A Laboratory Manual.* 2ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. / Samson M, et al. (1996) *Nature*, 382(6593):722-725. / Samulski et al. (1989) *J. Virol.* 63:3822-3828. / Sanchez-Pescador R. (1988) *J. Clin. Microbiol.* 26(10):1934-1938. / Sarkar, G. and Sommer S.S. (1991) *Biotechniques.* / Sauer B. et al. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:5166-5170. / Schedl A. et al., 1993a, Nature, 362: 258-261. / Schedl et al., 1993b, Nucleic Acids Res., 21: 4783-4787. / Schena et al. *Proc. Natl. Acad. Sci. U.S.A.* 93:10614-10619, 1996 / Schena et al. *Science* 270:467-470, 1995 / Schneider et al.(1997) *Arlequin: A Software For Population Genetics Data Analysis.* University of Geneva. / Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation / Sczakiel G. et al. (1995) *Trends Microbiol.* 3(6):213-217. / Shay J.W. et al., 1991, Biochem. Biophys. Acta, 1072: 1-7. / Sheffield, V.C. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 49:699-706. / Shizuya et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797. / Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW, *Nat Genet* 1996;14(4):450-456 / Smith (1957) *Ann. Hum. Genet.* 21:254-276. / Smith et al. (1983) *Mol. Cell. Biol.* 3:2156-2165. / Sosnowski RG, Tu E, Butler WF, O'Connell JP, Heller MJ, *Proc Natl Acad Sci U S A* 1997;94(4):1119-1123 / Sternberg N.L. (1992) *Trends Genet.* 8:1-16. / Sternberg N.L. (1994) *Mamm. Genome.* 5:397-404. / Stryer, L., *Biochemistry,* 4th edition, 1995 / Syvanen AC, *Clin Chim Acta* 1994;226(2):225-236 / Tacson et al. (1996) *Nature Medicine.* 2(8):888-892. / Te Ricle et al. (1990) Nature. 348:649-651. / Thomas K.R. et al. (1986) *Cell.* 44:419-428. / Thomas K.R. et al. (1987) *Cell.* 51:503-512. / Thompson et al., 1994, Nucleic Acids Res. 22(2):4673-4680 / Tsuzuki, T. and Rancourt, D.E. (1998) Nucleic Acids Res., 26(4):988-993 / Tur-Kaspa et al. (1986) *Mol. Cell. Biol.* 6:716-718. / Tyagi et al. (1998) *Nature Biotechnology.* 16:49-53. / Urdea M.S. (1988) *Nucleic Acids Research.* 11:4937-4957. / Urdea M.S. et al.(1991) *Nucleic Acids Symp. Ser.* 24:197-200. / Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33:988-991 (1971) / Van der Lugt et al. (1991) *Gene.* 105:263-267. / Vlasak R. et al. (1983) *Eur. J. Biochem.* 135:123-126. / Wabiko et al. (1986) *DNA.*5(4):305-314. / Walker et al. (1996) *Clin. Chem.* 42:9-13. / Weir, B.S. (1996) *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, U.S.A.* / White, M.B. et al. (1992) *Genomics.* 12:301-306. / White, M.B. et al. (1997) *Genomics.* 12:301-306. / Wong et al. (1980) *Gene.* 10:87-94. / Wood S.A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 4582-4585. / Wu and Wu (1987) *J. Biol. Chem.* 262:4429-4432. / Wu and Wu (1988) *Biochemistry.* 27:887-892. / Wu et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:2757. / Yagi T. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:9918-9922. / Zou Y. R. et al. (1994) *Curr. Biol.* 4:1099-1103.

SEQUENCE LISTING FREE TEXT

The following free text appears in the accompanying Sequence Listing :

Potential 5' regulatory region

Stop

Homology with sequence in ref

Diverging nucleotide

Deletion of a

In ref

Polymorphic base

Or

Potential

Complement

Probe

Upstream amplification primer

Downstream amplification primer

Polymorphic amino acid Val or Ile

Sequencing oligonucleotide primerPU

Sequencing oligonucleotide primer RP

CLAIMS

1. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069.

2. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises a C at position 16348 of SEQ ID No 1.

3. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329.

4. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complementary sequence thereof, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1.

5. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 2 or the complementary sequence thereof, wherein said contiguous span comprises a T at position 197, an A at position 453, or a G at position 779 of SEQ ID No 2.

6. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID 1 or the complement thereof, wherein said span includes a *FLAP*-related biallelic marker in said sequence.

7. A polynucleotide according to claim 6, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A10, A22 to A28, and the complements thereof.

8. A polynucleotide according to claim 6, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A11 to A13, A15, A17 to A21, and the complements thereof.

9. A polynucleotide according to claim 6, wherein said *FLAP*-related biallelic marker is selected from the group consisting of A14, A16, and the complements thereof.

10. A polynucleotide according to any one of claims 6 to 9, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

11. A polynucleotide according to claim 10, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide.

12. A polynucleotide according to claim 10, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: P1 to P28, and the complementary sequences thereto.

13. A polynucleotide according to any one of claims 1 to 9, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.

14. A polynucleotide according to any one of claims 6 to 9, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

15. A polynucleotide according to any one of claims 6 to 9, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of a *FLAP*-related biallelic marker in said sequence.

16. A polynucleotide according to claim 15, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said *FLAP*-related biallelic marker in said sequence.

17. A polynucleotide according to claim 16, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D28, and E1 to E28.

18. An isolated, purified, or recombinant polynucleotide consisting essentially of a sequence selected from the following sequences: B1 to B17 and C1 to C17.

19. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 in SEQ ID No 3.

20. A polynucleotide for use in a hybridization assay for determining the identity of the nucleotide at a *FLAP*-related biallelic marker in SEQ ID No1 or the complement thereof.

21. A polynucleotide for use in a sequencing assay for determining the identity of the nucleotide at a *FLAP*-related biallelic marker in SEQ ID No:1 or the complement thereof.

22. A polynucleotide for use in an enzyme-based mismatch detection assay for determining the identity of the nucleotide at a *FLAP*-related biallelic marker in SEQ ID No 1 or the complement thereof.

23. A polynucleotide for use in amplifying a segment of nucleotides comprising a *FLAP*-related biallelic marker in SEQ ID No 1 or the complement thereof.

24. A polynucleotide according to any one of claims 1 to 23 attached to a solid support.

25. An array of polynucleotides comprising at least one polynucleotide according to claim 24.

26. An array according to claim 25, wherein said array is addressable.

27. A polynucleotide according to any one of claims 1 to 23 further comprising a label.

28. A recombinant vector comprising a polynucleotide according to any one of claims 1, 2, 3, 4, 5, and 19.

29. A host cell comprising a recombinant vector according to claim 28.

30. A non-human host animal or mammal comprising a recombinant vector according to claim 28.

31. A mammalian host cell comprising a *FLAP* gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1, 2, 3, 4, 5, and 19.

32. A non-human host mammal comprising a *FLAP* gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1, 2, 3, 4, 5, and 19.

33. A method of genotyping comprising determining the identity of a nucleotide at a *FLAP*-related biallelic marker of SEQ ID No 1 or the complement thereof in a biological sample.

34. A method according to claim 33, wherein said biological sample is derived from a single subject.

35. A method according to claim 34, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome.

36. A method according to claim 33, wherein said biological sample is derived from multiple subjects.

37. A method according to claim 33, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

38. A method according to claim 37, wherein said amplifying is performed by PCR.

39. A method according to claim 33, wherein said determining is performed by a hybridization assay.

40. A method according to claim 33, wherein said determining is performed by a sequencing assay.

41. A method according to claim 33, wherein said determining is performed by a microsequencing assay.

42. A method according to claim 33, wherein said determining is performed by an enzyme-based mismatch detection assay.

43. A method of estimating the frequency of an allele of a *FLAP*-related biallelic marker in a population comprising:

a) genotyping individuals from said population for said biallelic marker according to the method of claim 33; and

b) determining the proportional representation of said biallelic marker in said population..

44. A method of detecting an association between a genotype and a trait, comprising the steps of:

a) determining the frequency of at least one *FLAP*-related biallelic marker in trait positive population according to the method of claim 43;

b) determining the frequency of at least one *FLAP*-related biallelic marker in a control population according to the method of claim 43; and

c) determining whether a statistically significant association exists between said genotype and said trait.

45. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

a) genotyping at least one *FLAP*-related biallelic marker according to claim 34 for each individual in said population;

b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and

c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

46. A method according to claim 45, wherein said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark algorithm, or an expectation-maximization algorithm.

47. A method of detecting an association between a haplotype and a trait, comprising the steps of:

a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 45;

b) estimating the frequency of said haplotype in a control population according to the method of claim 45; and

c) determining whether a statistically significant association exists between said haplotype and said trait.

48. A method according to claim 44, wherein said genotyping steps a) and b) are performed on a single pooled biological sample derived from each of said populations.

49. A method according to claim 44, wherein said genotyping steps a) and b) performed separately on biological samples derived from each individual in said populations.

50. A method according to either claim 44 or 47, wherein said trait is a disease involving the leukotriene pathway, a beneficial response to treatment with agents acting on the leukotriene pathway or side-effects related to treatment with agents acting on the leukotriene pathway.

51. A method according to claim 50 wherein said trait is asthma.

52. A method according to either claim 44 or 47, wherein said control population is a trait negative population.

53. A method according to either claim 44 or 47, wherein said case control population is a random population.

54. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 of SEQ ID No 3.

55. An isolated or purified antibody composition are capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 54, wherein said epitope comprises an isoleucine residue at amino acid position 127 in SEQ ID No 3.

56. A method of determining whether an individual is at risk of developing asthma, comprising:

a) genotyping at least one *FLAP*-related biallelic marker according to the method of claim 35; and

b) correlating the result of step a) with a risk of developing asthma.

57. A method according to any one of claims 33, 48, 44, 45, 47, and 56 wherein said *FLAP*-related biallelic marker is selected from the group consisting of A1 to A28.

58. A method according to claim 57, wherein said *FLAP*-related biallelic marker is selected from the following list of biallelic markers: A2, A14, A16, A18, A19, A22, and A23.

59. A method according to claim 57, wherein said *FLAP*-related biallelic marker is the biallelic marker A19.

60. A diagnostic kit comprising a polynucleotide according to any one of claims 6 to 19, 24 and 27.

61. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a C at position 16348 of SEQ ID No 1;

c) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329;

d) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1;

e) a contiguous span of at least 12 nucleotides of SEQ ID No 2, wherein said contiguous span comprises a T at position 197, an A at position 453, or a G at position 779 of SEQ ID No 2; and

f) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), or e).

62. A computer readable medium having stored thereon a sequence consisting of a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 of SEQ ID No 3.

63. A computer system comprising a processor and a data storage device wherein said data storage device a computer readable medium according to with claim 61 or 62.

64. A computer system according to claim 63, further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

65. A computer system of Claim 64 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

66. A computer system of Claim 63 further comprising an identifier which identifies features in said sequence.

67. A method for comparing a first sequence to a reference sequence, comprising the steps of:

reading said first sequence and said reference sequence through use of a computer program which compares sequences; and

determining differences between said first sequence and said reference sequence with said computer program,

wherein said first sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a C at position 16348 of SEQ ID No 1;

c) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329;

d) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1;

e) a contiguous span of at least 12 nucleotides of SEQ ID No 2, wherein said contiguous span comprises a T at position 197, an A at position 453, or a G at position 779 of SEQ ID No 2; and

f) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), or e); and

a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 of SEQ ID No 3.

68. A method according to Claim 6, wherein said step of determining differences between the first sequence and the reference sequence comprises identifying at least one polymorphism.

69. A method for identifying a feature in a sequence, comprising the steps of:

reading said sequence through the use of a computer program which identifies features in sequences; and

identifying features in said sequence with said computer program;

wherein said sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 1: 1-7007, 8117-15994, 16550-24058, 24598-27872, 28413-35976, and 36927-43069;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a C at position 16348 of SEQ ID No 1;

c) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises the following nucleotide positions of SEQ ID No 1: 7612-7637, 24060-24061, 24067-24068, 27903-27905, and 28327-28329;

d) a contiguous span of at least 12 nucleotides of SEQ ID No 1, wherein said contiguous span comprises a nucleotide selected from the group consisting of an A at position 7445, an A at position 7870, a T at position 16288, an A at position 16383, a T at position 24361, a G at position 28336, a T at position 28368, an A at position 36183,and a G at position 36509 of SEQ ID No 1;

e) a contiguous span of at least 12 nucleotides of SEQ ID No 2, wherein said contiguous span comprises a T at position 197, an A at position 453, or a G at position 779 of SEQ ID No 2; and

f) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), or e); and

a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 3, wherein said contiguous span includes an isoleucine residue at amino acid position 127 of SEQ ID No 3.

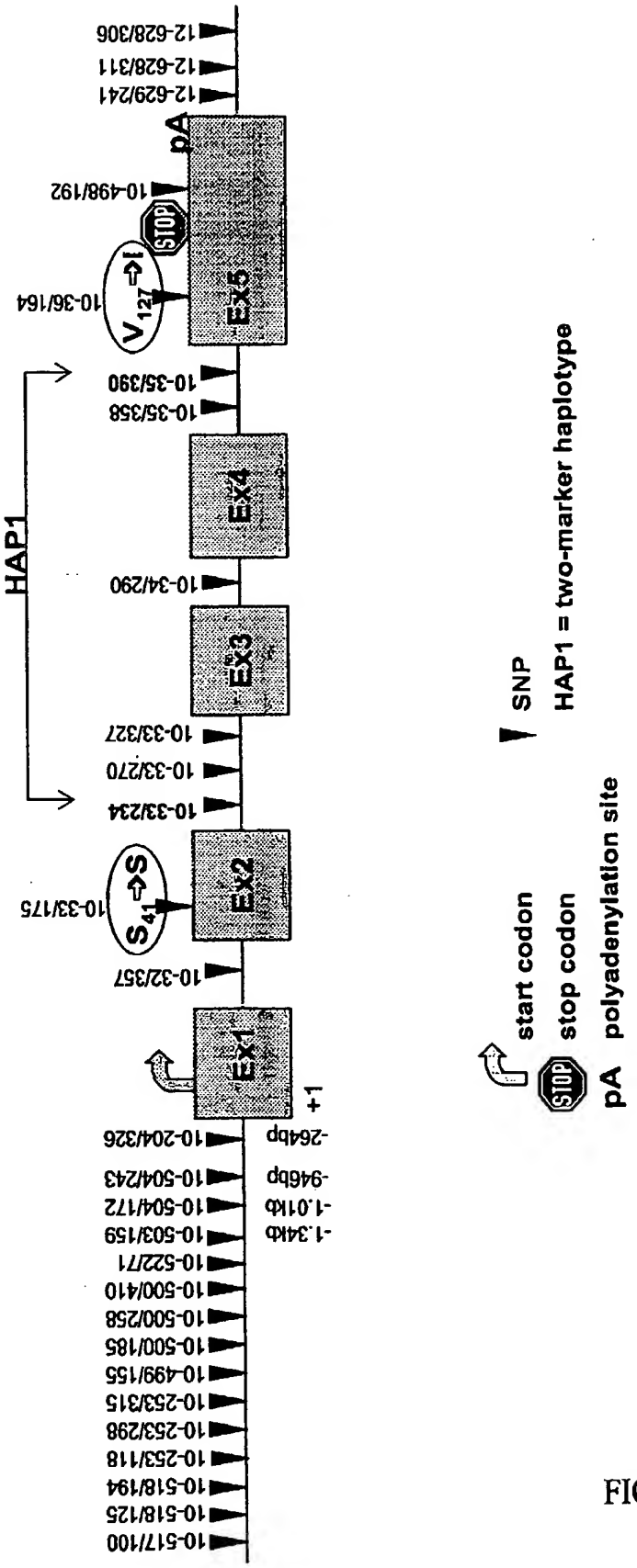
12-628/306
12-628/311
12-629/241
pA
10-498/192
STOP
10-36/164
V_{127}→I
EX5
10-35/390
10-35/358
EX4
HAP1
10-34/290
EX3
10-33/327
10-33/270
10-33/234
S_{41}→S
10-33/175
EX2
10-32/357
EX1
+1
-264bp
10-204/326
-946bp
10-504/243
-1.01kb
10-504/172
-1.34kb
10-503/159
10-522/71
10-500/410
10-500/258
10-500/185
10-499/155
10-253/315
10-253/298
10-253/118
10-518/194
10-518/125
10-517/100
SNP
HAP1 = two-marker haplotype
start codon
stop codon
STOP
pA polyadenylation site

FIGURE 1

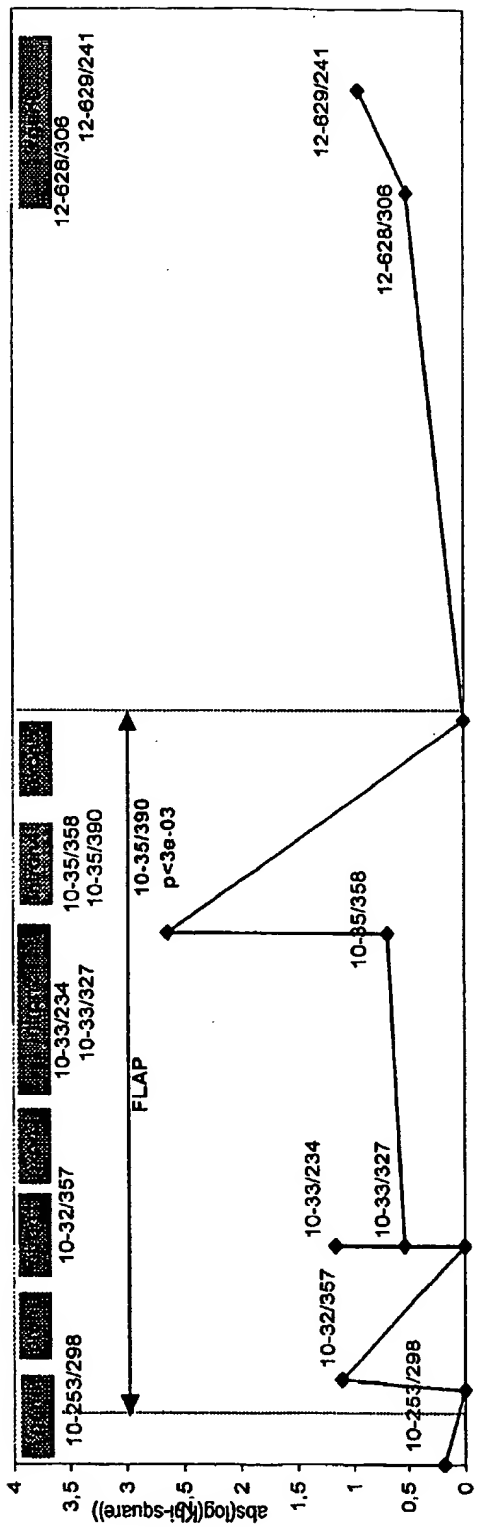
abs(log(Khi-square))
4
3,5
3
2,5
2
1,5
1
0,5
0
10-253/298
10-32/357
10-33/234
10-33/327
10-35/358
10-35/390
12-628/306
12-629/241
FLAP
10-35/390
p<3e-03

FIGURE 2

MARKERS	10-253/298	10-33/175	10-33/234	10-33/327	10-35/358	10-35/390	12-628/306	12-629/241	ESTIMATED FREQUENCIES			
FLAP	5'gene	ex2	In2		In4		3'gene		Frequencies haplotype		Odds ratio	Pvalue (1df)
cases/controls	287/186	295/174	295/274	295/270	291/280	295/272	284/185	283/182	cases	controls		
freq % case/controls	95/95 (C)	99/98 (C)	49/44 (A)	78/76 (T)	72/69 (G)	31/23 (C)	88/90 (C)	76/72 (G)				
diff.freq.all. (cases - controls)	0,5	1,8	5,3	2,6	3,4	9	2,1	4,6				
pvalue	6,55E-01	1,35E-02	6,93E-02	2,94E-01	2,08E-01	2,29E-03	3,17E-01	1,14E-01				
HAP 1	293 vs 265		A			T			0.283	0.197	1.81	(8.2e-04)
HAP 2	281 vs 177		A					G	0.305	0.210	1.85	(1.6e-03)
HAP 3	293 vs 261			T		T			0.307	0.224	1.53	(1.8e-03)
HAP 4	289 vs 271				G	T			0.304	0.231	1.46	(5.2e-03)
HAP 5	293 vs 168	C				T			0.309	0.226	1.53	(6.9e-03)
HAP 6	293 vs 265		A	T					0.276	0.208	1.46	(7.3e-03)
HAP 7	282 vs 178					T		G	0.314	0.233	1.50	(7.7e-03)
HAP 37	281 vs 176		A			T	C		0.265	0.171	1.76	(8.6e-04)
HAP 38	280 vs 173		A			T		G	0.292	0.194	1.71	(1.0e-03)
HAP 39	289 vs 264		A		G	T			0.283	0.199	1.59	(1.1e-03)
HAP 40	278 vs 175		A				C	G	0.271	0.180	1.70	(1.6e-03)
HAP 41	284 vs 176	C	A			T			0.287	0.195	1.66	(1.7e-03)
HAP 121	277 vs 171		A			T	C	G	0.265	0.169	1.77	(8.6e-04)
HAP 122	278 vs 173		A		G	T		G	0.290	0.195	1.69	(1.3e-03)
HAP 123	279 vs 176		A		G	T	C		0.264	0.175	1.70	(1.7e-03)
HAP 124	276 vs 175		A		G		C	G	0.271	0.181	1.69	(1.7e-03)
HAP 125	280 vs 174	C	A			T	C		0.265	0.176	1.69	(1.8e-03)
HAP 247	275 vs 171		A		G	T	C	G	0.265	0.170	1.77	(8.1e-04)
HAP 248	276 vs 169	C	A			T	C	G	0.265	0.172	1.74	(1.3e-03)
HAP 373	274 vs 169	C	A		G	T	C	G	0.265	0.172	1.73	(1.4e-03)
HAP 457	273 vs 163	C	A	T	G	T	C	G	0.247	0.167	1.64	(5.2e-03)

FIGURE 3

Markers	10-33/234			10-35/390		
	In2			In4		
ALT vs US	A			T		
cases vs US controls	5.3 (51 vs 56)	6,93E-02		9 (31 vs 23)	2,29E-03	
cases vs FRENCH controls	7.4 (49 vs 42)	3,39E-02		9.2 (31 vs 22)	1,75E-03	
ASSOCIATION	diff all. Freq	pvalue		diff all. Freq	pvalue	

HAPLOTYPE (AT)	sample sizes cases vs controls	haplotype frequencies		p-excess	odds-ratio	P value	PERMUTATIONS TEST RESULTS		
		cases	controls				Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
Asthmatics vs US controls	293 vs 265	0,283	0,197	10,7	1,61	8,20E-04	1,2	7,4	0/1000
							1,2	12,9	1/10 000
Asthmatics vs FRENCH controls	293 vs 154	0,283	0,192	11,33	1,67	2,70E-03	1,1	8,9	0/1000

FIGURE 4

<110> Genset SA
<120> Genomic Sequence Of The 5-Lipoxygenase-Activating Protein (FLAP), Polymorphic Markers Thereof And Methods For Detection Of Asthma.
<130> FLAP
<150> 60/081,893
<151> 1998-04-15
<150> 60/091,314
<151> 1998-06-30
<150> 60/123,406
<151> 1999-03-08
<160> 15
<170> Patent.pm

<210> 1
<211> 43069
<212> DNA
<213> homo sapiens

<220>
<221> misc_feature
<222> 1..7708
<223> potential 5'regulatory region

<220>
<221> misc_feature
<222> 36604..43069
<223> potential 3'regulatory region

<220>
<221> exon
<222> 7709..7852
<223> exon1

<220>
<221> exon
<222> 16236..16335
<223> exon2

<220>
<221> exon
<222> 24227..24297
<223> exon3

<220>
<221> exon
<222> 28133..28214
<223> exon4

<220>
<221> exon
<222> 36128..36605
<223> exon5

<220>
<221> misc_feature
<222> 7783..7785
<223> ATG

<220>
<221> misc_feature
<222> 36288..36290
<223> stop : TAA

<220>
<221> polyA_signal
<222> 36581..36586
<223> AATAAA

<220>
<221> misc_feature
<222> 7008..8116
<223> homology with sequence in ref genbank : M60470

<220>
<221> misc_feature
<222> 15995..16549
<223> homology with sequence in ref genbank : M63259

<220>
<221> misc_feature
<222> 24059..24597
<223> homology with sequence in ref genbank : M63260

<220>
<221> misc_feature
<222> 27873..28412
<223> homology with sequence in ref genbank : M63261

<220>
<221> misc_feature
<222> 35977..36926
<223> homology with sequence in ref genbank : M63262

<220>
<221> misc_feature
<222> 7613
<223> diverging nucleotide deletion of a A in ref : M60470

<220>
<221> misc_feature
<222> 16347
<223> diverging nucleotide G in ref : M63259

<220>
<221> misc_feature
<222> 16348
<223> diverging nucleotide A in ref : M63259

<220>
<221> misc_feature
<222> 24060
<223> diverging nucleotide deletion of a G in ref : M63260

<220>
<221> misc_feature
<222> 24067
<223> diverging nucleotide deletion of a G in ref : M63260

<220>
<221> misc_feature
<222> 27903
<223> diverging nucleotide deletion of a C in ref : M63261

<220>
<221> misc_feature

<222> 28327
<223> diverging nucleotide deletion of a G in ref : M63261

<220>
<221> misc_feature
<222> 3851..4189
<223> 10-517

<220>
<221> misc_feature
<222> 4120..4390
<223> 10-518

<220>
<221> misc_feature
<222> 4373..4792
<223> 10-253

<220>
<221> misc_feature
<222> 4814..5043
<223> 10-499

<220>
<221> misc_feature
<222> 4956..5422
<223> 10-500

<220>
<221> misc_feature
<222> 5524..5996
<223> 10-522

<220>
<221> misc_feature
<222> 6218..6672
<223> 10-503

<220>
<221> misc_feature
<222> 6522..6790
<223> 10-504

<220>
<221> misc_feature
<222> 7120..7574
<223> 10-204

<220>
<221> misc_feature
<222> 7513..7933
<223> 10-32

<220>
<221> misc_feature
<222> 16114..16533
<223> 10-33

<220>
<221> misc_feature
<222> 24072..24425
<223> 10-34

<220>
<221> misc_feature
<222> 27978..28401
<223> 10-35

<220>
<221> misc_feature
<222> 36020..36465
<223> 10-36

<220>
<221> misc_feature
<222> 36318..36669
<223> 10-498

<220>
<221> misc_feature
<222> 38441..38840
<223> 12-629

<220>
<221> misc_feature
<222> 42233..42749
<223> complement 12-628

<220>
<221> allele
<222> 3950
<223> 10-517-100 : polymorphic base G or C

<220>
<221> allele
<222> 4243
<223> 10-518-125 : polymorphic base G or T

<220>
<221> allele
<222> 4312
<223> 10-518-194 : polymorphic base A or G

<220>
<221> allele
<222> 4490
<223> 10-253-118 : polymorphic base A or G

<220>
<221> allele
<222> 4670
<223> 10-253-298 : polymorphic base G or C

<220>
<221> allele
<222> 4687
<223> 10-253-315 : polymorphic base C or T

<220>
<221> allele
<222> 4968
<223> 10-499-155 : polymorphic base A or G

<220>
<221> allele
<222> 5140

<223> 10-500-185 : polymorphic base C or T

<220>
<221> allele
<222> 5213
<223> 10-500-258 : polymorphic base G or T

<220>
<221> allele
<222> 5364
<223> 10-500-410 : polymorphic base A or G

<220>
<221> allele
<222> 5594
<223> 10-522-71 : polymorphic base A or G

<220>
<221> allele
<222> 6370
<223> 10-503-159 : polymorphic base G or T

<220>
<221> allele
<222> 6693
<223> 10-504-172 : polymorphic base A or T

<220>
<221> allele
<222> 6763
<223> 10-504-243 : polymorphic base A or C

<220>
<221> allele
<222> 7445
<223> 10-204-326 : polymorphic base A or G

<220>
<221> allele
<222> 7870
<223> 10-32-357 : polymorphic base A or C

<220>
<221> allele
<222> 16288
<223> 10-33-175 : polymorphic base C or T

<220>
<221> allele
<222> 16347
<223> 10-33-234 : polymorphic base A or C

<220>
<221> allele
<222> 16383
<223> 10-33-270 : polymorphic base A or G

<220>
<221> allele
<222> 16440
<223> 10-33-327 : polymorphic base C or T

<220>

<221> allele
<222> 24361
<223> 10-34-290 : polymorphic base G or T

<220>
<221> allele
<222> 28336
<223> 10-35-358 : polymorphic base G or C

<220>
<221> allele
<222> 28368
<223> 10-35-390 : polymorphic base C or T

<220>
<221> allele
<222> 36183
<223> 10-36-164 : polymorphic base A or G

<220>
<221> allele
<222> 36509
<223> 10-498-192 : polymorphic base A or G

<220>
<221> allele
<222> 38681
<223> 12-629-241 : polymorphic base G or C

<220>
<221> allele
<222> 42440
<223> 12-628-311 : polymorphic base T or C

<220>
<221> allele
<222> 42445
<223> 12-628-306 : polymorphic base G or A

<220>
<221> misc_binding
<222> 3930..3949
<223> 10-517-100.mis1 potential

<220>
<221> misc_binding
<222> 3951..3970
<223> complement 10-517-100.mis2 potential

<220>
<221> misc_binding
<222> 4223..4242
<223> 10-518-125.mis1 potential

<220>
<221> misc_binding
<222> 4244..4263
<223> complement 10-518-125.mis2 potential

<220>
<221> misc_binding
<222> 4292..4311
<223> 10-518-194.mis1 potential

<220>
<221> misc_binding
<222> 4313..4332
<223> complement 10-518-194.mis2 potential

<220>
<221> misc_binding
<222> 4470..4489
<223> 10-253-118.mis1 potential

<220>
<221> misc_binding
<222> 4491..4510
<223> complement 10-253-118.mis2 potential

<220>
<221> misc_binding
<222> 4650..4669
<223> 10-253-298.mis1

<220>
<221> misc_binding
<222> 4671..4690
<223> complement 10-253-298.mis2 potential

<220>
<221> misc_binding
<222> 4667..4686
<223> 10-253-315.mis1 potential

<220>
<221> misc_binding
<222> 4688..4707
<223> complement 10-253-315.mis2 potential

<220>
<221> misc_binding
<222> 4948..4967
<223> 10-499-155.mis1 potential

<220>

<221> misc_binding
<222> 4969..4988
<223> complement 10-499-155.mis2 potential

<220>
<221> misc_binding
<222> 5120..5139
<223> 10-500-185.mis1 potential

<220>
<221> misc_binding
<222> 5141..5160
<223> complement 10-500-185.mis2 potential

<220>
<221> misc_binding
<222> 5193..5212
<223> 10-500-258.mis1 potential

<220>

<221> misc_binding
<222> 5214..5233
<223> complement 10-500-258.mis2 potential

<220>
<221> misc_binding
<222> 5344..5363
<223> 10-500-410.mis1 potential

<220>
<221> misc_binding
<222> 5365..5384
<223> complement 10-500-410.mis2 potential

<220>
<221> misc_binding
<222> 5574..5593
<223> 10-522-71.mis1 potential

<220>
<221> misc_binding
<222> 5595..5614
<223> complement 10-522-71.mis2 potential

<220>
<221> misc_binding
<222> 6350..6369
<223> 10-503-159.mis1 potential

<220>
<221> misc_binding
<222> 6371..6390
<223> complement 10-503-159.mis2 potential

<220>
<221> misc_binding
<222> 6673..6692
<223> 10-504-172.mis1 potential

<220>
<221> misc_binding
<222> 6694..6713
<223> complement 10-504-172.mis2 potential

<220>
<221> misc_binding
<222> 6743..6762
<223> 10-504-243.mis1 potential

<220>
<221> misc_binding
<222> 6764..6783
<223> complement 10-504-243.mis2 potential

<220>
<221> misc_binding
<222> 7425..7444
<223> 10-204-326.mis1 potential

<220>
<221> misc_binding
<222> 7446..7465
<223> complement 10-204-326.mis2

<220>
<221> misc_binding
<222> 7850..7869
<223> 10-32-357.mis1

<220>
<221> misc_binding
<222> 7871..7890
<223> complement 10-32-357.mis2 potential

<220>
<221> misc_binding
<222> 16268..16287
<223> 10-33-175.mis1

<220>
<221> misc_binding
<222> 16289..16308
<223> complement 10-33-175.mis2 potential

<220>
<221> misc_binding
<222> 16327..16346
<223> 10-33-234.mis1

<220>
<221> misc_binding
<222> 16348..16367
<223> complement 10-33-234.mis2 potential

<220>
<221> misc_binding
<222> 16363..16382
<223> 10-33-270.mis1

<220>
<221> misc_binding
<222> 16384..16403
<223> complement 10-33-270.mis2 potential

<220>
<221> misc_binding
<222> 16420..16439
<223> 10-33-327.mis1

<220>
<221> misc_binding
<222> 16441..16460
<223> complement 10-33-327.mis2 potential

<220>
<221> misc_binding
<222> 24341..24360
<223> 10-34-290.mis1

<220>
<221> misc_binding
<222> 24362..24381
<223> complement 10-34-290.mis2 potential

<220>
<221> misc_binding

<222> 28316..28335
<223> 10-35-358.mis1 potential

<220>
<221> misc_binding
<222> 28337..28356
<223> complement 10-35-358.mis2

<220>
<221> misc_binding
<222> 28348..28367
<223> 10-35-390.mis1 potential

<220>
<221> misc_binding
<222> 28369..28388
<223> complement 10-35-390.mis2 potential

<220>
<221> misc_binding
<222> 36163..36182
<223> 10-36-164.mis1 potential

<220>
<221> misc_binding
<222> 36184..36203
<223> complement 10-36-164.mis2

<220>
<221> misc_binding
<222> 36489..36508
<223> 10-498-192.mis1 potential

<220>
<221> misc_binding
<222> 36510..36529
<223> complement 10-498-192.mis2 potential

<220>
<221> misc_binding
<222> 38661..38680
<223> 12-629-241.mis1

<220>
<221> misc_binding
<222> 38682..38701
<223> complement 12-629-241.mis2 potential

<220>
<221> misc_binding
<222> 42420..42439
<223> 12-628-311.mis2 potential

<220>
<221> misc_binding
<222> 42441..42460
<223> complement 12-628-311.mis1 potential

<220>
<221> misc_binding
<222> 42425..42444
<223> 12-628-306.mis2 potential

<220>
<221> misc_binding
<222> 42446..42465
<223> complement 12-628-306.misl

<220>
<221> misc_binding
<222> 3927..3973
<223> 10-517-100.probe potential

<220>
<221> misc_binding
<222> 4220..4266
<223> 10-518-125.probe potential

<220>
<221> misc_binding
<222> 4289..4335
<223> 10-518-194.probe potential

<220>
<221> misc_binding
<222> 4467..4513
<223> 10-253-118.probe potential

<220>
<221> misc_binding
<222> 4647..4693
<223> 10-253-298.probe potential

<220>
<221> misc_binding
<222> 4664..4710
<223> 10-253-315.probe potential

<220>
<221> misc_binding
<222> 4945..4991
<223> 10-499-155.probe potential

<220>
<221> misc_binding
<222> 5117..5163
<223> 10-500-185.probe potential

<220>
<221> misc_binding
<222> 5190..5236
<223> 10-500-258.probe potential

<220>
<221> misc_binding
<222> 5341..5387
<223> 10-500-410.probe potential

<220>
<221> misc_binding
<222> 5571..5617
<223> 10-522-71.probe potential

<220>
<221> misc_binding
<222> 6347..6393

<223> 10-503-159.probe potential

<220>
<221> misc_binding
<222> 6670..6716
<223> 10-504-172.probe potential

<220>
<221> misc_binding
<222> 6740..6786
<223> 10-504-243.probe potential

<220>
<221> misc_binding
<222> 7422..7468
<223> 10-204-326.probe potential

<220>
<221> misc_binding
<222> 7847..7893
<223> 10-32-357.probe potential

<220>
<221> misc_binding
<222> 16265..16311
<223> 10-33-175.probe potential

<220>
<221> misc_binding
<222> 16324..16370
<223> 10-33-234.probe potential

<220>
<221> misc_binding
<222> 16360..16406
<223> 10-33-270.probe potential

<220>
<221> misc_binding
<222> 16417..16463
<223> 10-33-327.probe potential

<220>
<221> misc_binding
<222> 24338..24384
<223> 10-34-290.probe potential

<220>
<221> misc_binding
<222> 28313..28359
<223> 10-35-358.probe potential

<220>
<221> misc_binding
<222> 28345..28391
<223> 10-35-390.probe potential

<220>
<221> misc_binding
<222> 36160..36206
<223> 10-36-164.probe potential

<220>

<221> misc_binding
<222> 36486..36532
<223> 10-498-192.probe potential

<220>
<221> misc_binding
<222> 38658..38704
<223> 12-629-241.probe potential

<220>
<221> misc_binding
<222> 42417..42463
<223> 12-628-311.probe potential

<220>
<221> misc_binding
<222> 42422..42468
<223> 12-628-306.probe potential

<220>
<221> primer_bind
<222> 3851..3869
<223> upstream amplification primer 10-517

<220>
<221> primer_bind
<222> 4171..4189
<223> downstream amplification primer 10-517 , complement

<220>
<221> primer_bind
<222> 4120..4138
<223> upstream amplification primer 10-518

<220>
<221> primer_bind
<222> 4372..4390
<223> downstream amplification primer 10-518 , complement

<220>
<221> primer_bind
<222> 4373..4391
<223> upstream amplification primer 10-253

<220>
<221> primer_bind
<222> 4773..4792
<223> downstream amplification primer 10-253 , complement

<220>
<221> primer_bind
<222> 4814..4833
<223> upstream amplification primer 10-499

<220>
<221> primer_bind
<222> 5026..5043
<223> downstream amplification primer 10-499 , complement

<220>
<221> primer_bind
<222> 4956..4972
<223> upstream amplification primer 10-500

<220>
<221> primer_bind
<222> 5405..5422
<223> downstream amplification primer 10-500 , complement

<220>
<221> primer_bind
<222> 5524..5542
<223> upstream amplification primer 10-522

<220>
<221> primer_bind
<222> 5978..5996
<223> downstream amplification primer 10-522 , complement

<220>
<221> primer_bind
<222> 6218..6235
<223> upstream amplification primer 10-503

<220>
<221> primer_bind
<222> 6652..6672
<223> downstream amplification primer 10-503 , complement

<220>
<221> primer_bind
<222> 6522..6539
<223> upstream amplification primer 10-504

<220>
<221> primer_bind
<222> 6772..6790
<223> downstream amplification primer 10-504 , complement

<220>
<221> primer_bind
<222> 7120..7137
<223> upstream amplification primer 10-204

<220>
<221> primer_bind
<222> 7557..7574
<223> downstream amplification primer 10-204 , complement

<220>
<221> primer_bind
<222> 7513..7531
<223> upstream amplification primer 10-32

<220>
<221> primer_bind
<222> 7914..7933
<223> downstream amplification primer 10-32 , complement

<220>
<221> primer_bind
<222> 16114..16132
<223> upstream amplification primer 10-33

<220>
<221> primer_bind

<222> 16515..16533
<223> downstream amplification primer 10-33 , complement

<220>
<221> primer_bind
<222> 24072..24089
<223> upstream amplification primer 10-34

<220>
<221> primer_bind
<222> 24408..24425
<223> downstream amplification primer 10-34 , complement

<220>
<221> primer_bind
<222> 27978..27995
<223> upstream amplification primer 10-35

<220>
<221> primer_bind
<222> 28384..28401
<223> downstream amplification primer 10-35 , complement

<220>
<221> primer_bind
<222> 36020..36039
<223> upstream amplification primer 10-36

<220>
<221> primer_bind
<222> 36446..36465
<223> downstream amplification primer 10-36 , complement

<220>
<221> primer_bind
<222> 36318..36337
<223> upstream amplification primer 10-498

<220>
<221> primer_bind
<222> 36652..36669
<223> downstream amplification primer 10-498 , complement

<220>
<221> primer_bind
<222> 38441..38460
<223> upstream amplification primer 12-629

<220>
<221> primer_bind
<222> 38820..38840
<223> downstream amplification primer 12-629 , complement

<220>
<221> primer_bind
<222> 42233..42253
<223> downstream amplification primer 12-628

<220>
<221> primer_bind
<222> 42731..42749
<223> upstream amplification primer 12-628 , complement

<400> 1

gtgtcagctc agtcttgcgg gttttgggtt gtccttgctt cccacacttc atgcctttct 60
ttccctcctg acagtctgcc ctttagattt taggattcag caccagccac agaaacagca 120
acctcactgt taagggttga attgtatctc cccaaaaggt aggttgaggc cctacctgcc 180
aggacttcag aatgtaacct catttgggaa tagcatcatt gcaaatataa ttaattaaga 240
tgagggcata ctggctcagg atgggctcct aattcaatac aactaatgtc cttctacgac 300
agccacagga agacagaaac gccaagggag aacaccatat gctgatggag gcagtggcag 360
ctgccagcca aggattataa ccagaagtca ggaaaaagca agaaggaatc ctcccttagt 420
gattttacag ggagcatagc cctgctgaca ccttgatttt ggacttttat tccccaaaac 480
tgtaaaacaa tacacttctg ttgttttaag ccactcagtt tgtgctactt tgttatggca 540
actccagaaa acaaaaatac actcagactg tttaatcaac ctccataatt gcataaggtc 600
taatccctat aataaatccc ttaaaaatgt ctgtgtatat gtatttaaaa atataaaata 660
tcttctagtg gttctgcatc tctggtcaat ccctgactga tacagaatat gtattttcat 720
ttctaatgat gaaatacctg aatgaaattt ctaggacata tggtaagtgt atgtttagct 780
tttaagaaac tgccaacttg ggggaattgc ttgaggccag gagttcaaac agcctgggta 840
acagtgatac cctgtctgta caaaataaaa aatattagca gcgtgtggtg gtgtgtgtct 900
gtagtcccag ctactcagga ggctgaggtg ggagattcac ctgagcccag atctttgaag 960
ttatagtgag ctatgatcac gccactgcac tctagcctgg gtgacagagt gagaaagctg 1020
gtctctaaaa aacaaacaaa caaaaaagaa actgtcaaac tcttcccaac atgttgccat 1080
ttttacattt accattttac attcttacca gcaatgattg atagttccag ttgctccata 1140
cccttgctga ccattccaat agatgtattg tgttatctca ttgtagttct aatttgtatt 1200
tccctagtga ttaatgatgt ttaacatctt ttcatgcacc tattggctat atgtatatct 1260
tctttagcaa aatatatgtt gttatttgaa gagcggaagt tttacatttt gatgaagtct 1320
aatttattga ttttttttttt cttagatggc tcatgctttt tgtgttatct aaaaaaaatt 1380
tgccttcttc atggtcacaa agactttctc ctatgttttc ttttggaagc tttatatttt 1440
tagtttttat gtttatgttt aagacccatt tctagttaca atttgtgtga ttttttggaa 1500
gggtcaaggt tcattttctt ttccataaga atgtacagtt gttctagcac ccttgttaaa 1560
aagactttcc tttccccatt gaactacttt gtcaaaaatc aactgagcat atatgggcat 1620
catgaatttt aatcctgtta gaactgaatg ttcccaaggc aggccatgcc catgactgac 1680
ctcctttcct tggattgcct acaaaacaga taaagctaag tctggagcaa agaaatccat 1740
gtctaaccta tatttttttt ttttttttttt tttagatggg gtctcgctct gtcacccagg 1800
ctggagtgca gtggcgtgat cccagctcac tgcaatctct gcctcctggg ttcaagtgat 1860
tctcctgcct cagcctcccg aggggctggg attgtaggcg tgcaccacta tgcccatcta 1920
atttttgtat ttttagtaga gatagggttt tgccattttg gccagactgt cttgaactcc 1980
tgacctcagg tgatctgcct gcctcggcct cccacagttt tgtgattata ggcatgagcc 2040
accgtgcccg gccttaacct ttgttttctt acacaacaca ctacgtgatg ttttccacat 2100
gcatgggtca tttgcttcat ttacgtacaa atgcataagc aatatactgt gtggtgtgag 2160
tttgtgatgg gaaaaggaag aggttttgcg gatactacac tggcttcctg ctatctgtct 2220
gtgtgaatgg ctatggactt tgtcttctat ttgttcgctt agcgcagata tgatcagctt 2280
acaacttaag attctagaga aagagggtca tatctgtaaa gcactctgag catgtgtgaa 2340
gtttaatcaa tagcatatga ggttacagca aattcactat ctttgtttct tcagctatag 2400
aatggcatga ggattcatct caatttagtt caattctgtt cagaaccatg agctagctgt 2460
tcatggaagg aaagcccacc tgattgtggc cagggaagga gaaacaacac tttaaccagg 2520
ttgatttggt tctcacagac accattggca tgtgacatct ggaacagacc atgcctggtc 2580
tctgttcgta tcacttacca ttcagctcaa tattggtctg aatattcttt agactgactg 2640
aaatgaaaag gaactgttgt gtaaccatcc ataattccag cctgtagacc tgggctgtat 2700
ctctatgccc tgcctggcac agaccccacc tcctgctcct tctccctcac caccagtcaa 2760
tccttgtcct aatgaacagg gagggcaacc ctgaatgggg agtggaggga agagatgtca 2820
tgagatggca acgtgcaccc tgaagtgagg atgaaggcta tgtgaatgtt gtaggctgac 2880
agccgggcat agtggccccg ttgccatggc gatggaggca tgttgatgcg aagtgtctgc 2940
acagctccta ggatttttaa cagcagctgg gcagagcctc ggcgtccctg aattgttgcc 3000
cccctgagtc actgcttggc cccagctgtc ctgatctctg ttgacaaatg gttgtccttc 3060
acagtcaaac tactaacagt actctaatta atgaatgtgc taattattct tgcctactcc 3120
cagcatattt gtctaactaa cctgtcacac acagatcagt gcagcatatg cataattacg 3180
gagagcgctg ggagcagggg atgggtggga gaggggtggg ctcgcagccc tgtcgctgtg 3240
ggatatttct tgtaaagtta cctttgctaa cggtcagatg tcgtggggat atgttatttc 3300
ccgtgaagtg tatatgtctt ccttttcttc ctttctaaga atctctcttc agggctgagg 3360
ggccattgct cagtgcttta gcctgtgagg ggattgccag gtacaaatgc agaaggacca 3420
gggagcccag gttctgaaga cgattccggt agcagcacgt agggtgatta aaactccaga 3480
ctttaaagcc agaccggcct gggcttgaac ccttgttctg ctccttgcta tgtgggtctt 3540
tgccttgacc acatttttttt ttttttttta agacaggatc tccctctctt gcccaggctg 3600
taatgcagtg ttgcgatcac agctcactga agcctccatc tctacagcct caagcgatcc 3660
tcctgcctca gccccgagta gctgggacta caggtctgtg ccaccacgtc cagctaattt 3720

actttttgtag agttgggggt cttgctatgt tgcccaggct gttctccaac tcctggactc 3780
aagccatcct ctagcctcgg ccttccaaag tgctgggact ataggcgtga gccacggtgc 3840
caggcccttg accacatttt taacccctct gaacctcagt ttcactttct gggcaatggg 3900
agggggggtaa tttgtccctc agagggttgc actgaggggc aaatgtgagg ctctgggtac 3960
aatgcccagt acagactagg tccccacgac acagccgctc agcggctccg gattctgggc 4020
tgctctggac tgcggccagg cggtcttctg cgggaatccg ggcaggcagg gcgggctgcg 4080
ctcccctccc cggctctccc ggtgcccctt gtctttttgt tctgtctcag cagctctcta 4140
ttaagatgaa tggcatttcc aaaggcttca cctctgataa gtgttcctct gcagctgcag 4200
ccagaatctt aatgtgcgcg ctgtaattta atggccgtct cggctattaa cacgctcttc 4260
tcgggtgaag tggactccct ccatccccgg gcctctgcac gtgctctgcg cactggctgg 4320
gggtgactcc aaggagctca gagcggggtg cccggcacct ctcgccaggc gcctttcgac 4380
cttctaaagc gcgaatggct ggacttttct cccatgtgtg gggccccaga aggtgtgggg 4440
ccccagaagg tgtggggtcc ctgcgttcca cggagcccgg aaggtttcca gtgatggtgg 4500
gggctgacca cgttggtccc cgtgggtgct gttttcatgt gccggcagat tgggatgagt 4560
ttaaaagaca gaagcgtgta ggatagagaa acttctttaa aaactggaaa ttttaatctg 4620
gggattataa ctattggaca gtcaagtgca agagtgaata cacttctcac tccctcctcc 4680
caatttttat ttgcgggatt agtcagtccc cctctgccac atgataattg tgagaactac 4740
cagggtcttc attctcctgc catctggttg acctctccaa gaatggacac ccgggcagcc 4800
tgggccaatg aggctgtcct aagagtttag atgagagaag tcagtctttg acaggtgatg 4860
gaagctgtaa aatgtaaaac tccacagttg gtgaagatgt ctccaggaaa caggtctgca 4920
gagagaatac gtttgacatg ctaagagaag ctgagagaga gcgagaggag agattggaag 4980
aaagacagag acagaggtag agagaaggga aagagagaga gaaagggaca gaagagagag 5040
aaaatagagg gggccgggcg cggtggctca cgcctgtaat ctcagcactt tgggaggccg 5100
aggcgggcag atcacgaggt caggagatcg agaccatccc ggctaacacg gtgaaacccc 5160
cgtctctact aaaaaatata aaaaaaatta gccaggcgtg gtggtgggtg cctgtagtcc 5220
cagctactga ggaggctgag acaggagaat ggcgtgaacc cgggaggcag agcttgcagt 5280
gagctgagat cgcgccactg cactccagcc tgggcaacag agcaagactc cgtctcaaaa 5340
aaaaaaaaaa aaaagagagg aagggcggga gagagagaga gagaaagctc tctagctcca 5400
aggcctaacc acatctctgt tcttttcaac ttcagctgtc agatttttag actctttgag 5460
tgaataaatt ctccttttg cttaaactag tttgagctaa gtttctattg cttgcaactg 5520
gaatactttg taagaggact ggccttcatt tctgatgcat tgtcactaag atgtaagtgt 5580
tagaagagct aacgctttat ggggttcaaa ctccttggct accaaaacct aaacatcccc 5640
tgaaacttac caaactgcag gtatgaattg gatctcacta aggtgaatat acaaatcttg 5700
caagtgctga gccctaacca atcttgtaat aactctgtgg tagttaattt tatgtcaaat 5760
tgattgagct aaaaaatgcc caggtagctg gtaaaatgtt tttttctggg tgtgttaggg 5820
agggtgtttc tgaaagagat cagcactgga atcagcggac taagtaaaga attcccaccc 5880
tcaccaatat ggtgggtgtc atcaatccac tgagggcctg aatagaacaa aaagcgggca 5940
gaagggcaaa ttccctcttc ttcttgagct gggccatcca tcttctcctg cccttggaca 6000
ctggagcccc ttgttctcca gcttttggat tcagactggg tcttgcacca ttgccctcca 6060
tcttctcctg cccttggaca ctggagcccc ttgttctcca gcttttggat tcagactggg 6120
tcttgcacca ttgccctcct tgatgctcag gcctttgaat gcagactggt ctccaccagc 6180
agcttttctg agtctccagc ttgcagatgg caaaccatga aacttcatgg tgtccatgag 6240
catgtgaacc aatttctatt ataaatctgc aatatatata tatgaggaga cttatttata 6300
tattggttca gtttctctgg agagccttgg ctaatataaa gtctatactc tacaaagtgc 6360
cctaggtact cagggagtac ccaagtgtgt catgaccagc ccgacagccc tggctgctgg 6420
cttccccgca cacaactctg cacgctgcct tcatcagcct ttctctctca gctgaaccga 6480
gggcattgaa gcgggcctct ggcactgtac ctatgaggga gcaatatctt cccctacact 6540
gacctcttcc gtgccgagat gcagccctcc ctgctgccac tagttacagt ggtccatgtt 6600
ccctttcaaa gtgaagtttt gataaaagca cctcttaacc aatgccaaat agctaagtct 6660
gggacaaaga ttgcaggtat tttgcatttt ccatgtaacc tcagagggat tgccattcac 6720
actgatctga gctgcagaat accaggcagc cacctcaccc acccagcagg tccactctta 6780
tactttctca gaaagcacag ccactctact cttattcagt tgaaaagaat ttccaggaag 6840
gtgtttctgc gattgcctca gaaaagtcag ttccctttgg gaatttccct tagggatcat 6900
ctgtaactcc atttctgcct tttacctgaa ttctttggtt tggtttgaat tctttggttt 6960
aatttatgaa ttcccttttat tacttttctc tgaagaaatg gagatatcag ctgtccctcc 7020
ccactgccat ttattccttc cttcattcaa accttatgtg gctgctactt accgtgtgtt 7080
aagtgttcac tttttttctt ggaattcaaa aaaagaagga cagtatttgg ggcacagatc 7140
ttttggtgtt ctatacattt ttttaaagtt tcattttaca tttgtgtgtg cgtgtgtgtg 7200
tgtgtgtgag acagtcttgc tctgttgccc aggctggagt gcagtggcat aatcattggc 7260
tcactgtagc ctcaaagtcc tgggcccaag cgatcttccc acctcagcca cccaaaatgc 7320
tggggttaca ggtttatgcc actctgtctg acctgaaagt tttgggttta ctttcccttc 7380
tttctctttg ctgaagtcag agatgatggc agcttccaga ttctctggtg cctgtgctgg 7440
gctcgtgctg gtcatggtct tgggtccagg attcattctg gagactctca gggaagtttc 7500

ccatgacaag gaaatgtagg agagtgtgct ggctttgcgt gctcctctgc caagccctgc 7560
ttctcctggt gggacacact gaaccacagc cagggcattt tggtggttag ttaaaaaaaa 7620
aaaaaaaaaa aaaaaaggaa gaagaaggca ctgtgtaatt gtgccgggga tcttcagaaa 7680
ttgtaatgat gaaagagtgc aagctctcac ttccccttcc tgtacagggc aggttgtgca 7740
gctggaggca gagcagtcct ctctggggag cctgaagcaa acatggatca agaaactgta 7800
ggcaatgttg tcctgttggc catcgtcacc ctcatcagcg tggtccagaa tggtaaggaa 7860
agcccttcac tcagggaaga acagaagggg agattttctt tgatggttgt ttggaagtca 7920
ggcttaaaca attgtgtctg tgtgtgcgca tgcacaaaca cttttacctt atctttattt 7980
tcttcttttt atttgaatgt atagggttgt gtgtatttct gtgtaaattt ggggttttcc 8040
tcctcttagt ctttcacttt tgtggtgatt accagtccca tttttagagc cagggctgca 8100
acttgaaggt tttgctaaaa ccctcaccga agtgtctatg atcagcattt taactattaa 8160
ttaatgtggc caggcaaggg gtggaaggtg agaagactag aaagggaaca tgatatacac 8220
atttactcag atactgggct tttctaacat ctgcagtgca attgaagtta ccagtcatct 8280
gcagtctaaa aagaaagtga ttttgggagg tgcgtagaaa aaatcatctt attatttttc 8340
ctctatatta ctttttttctt tttttctcct gaagaaactt tttttttgg tgataccttc 8400
tttttctcta gcacgtataa ttttggaagc atttttcata tgcagtgtat acttcagaaa 8460
gagagagaga gagaggaaaa ttgtcctgtt cagcgtttgc atttccatta ttcctgctat 8520
tagttaaaaa caacaacaac aacaaaaaac aagcaggata cctagatctg gaaaagggag 8580
aattgtgtag agctgtcttc ctaaagttct gagttagggc tgcctcagac cactttcata 8640
actatctcca gtggctttgt gttttatatt tattaagata gagaaaaaaa gagtaattac 8700
taagggcagc tgctgtagct ttatggtgat tactgaacat tgacatgctg tcacgttttt 8760
ggaactttga gtatttaatc actttgggat attctatttt cccccatctt gagtgtggac 8820
agatgctggt gatgtagcct tctgggcaca gagcaagcct ccccctcagc ctctgcacca 8880
gaaaggctca gcttcacaca ctccaagtat gttttctaca agaactacac tttgtggctt 8940
tctgacccaa acatttttat actaaattac acacaacaaa gttgtagctc agagagggaa 9000
caaatggctt atttaggcca ccattttctt gagccattat gatttcacac agggctccct 9060
tggccctgta aattggcaag gattccatta ttcaacccgc atacatgtac agagaccctg 9120
ctctggccca gatagtattc tgggtacagg cggatagagc aggaaacaaa acagctacag 9180
tgatggacag gtcagcctgc agcaatgcct gcagtctctg caaaggtagc tgtatgggtg 9240
ggcaggtggc tagcacttat tcagctctgg aaggatctcc cctctggcct ctcccctgac 9300
acccatcaat aaaactgagg agcatcggtg gacaggggac cttgtgcccc ctccctgcct 9360
gtgcagttgg ggctgaaccc agctacgaag tttgagctca ctctctccag ctccctctca 9420
attcagagct gaactgtggg aagcttcaga gctctctgtt tcaaggacag gttctcctca 9480
cctctcctaa tggaggtgca ccagggaact ggccctgctc tgcccagggc tttctcctgg 9540
actttgccat catggtctag caaaccctgt tcagattgag gtgagtggtg agatttcgaa 9600
ttctttttga cagataggat taagtcttct tctgtgggac aagtgggagg tagaggtaag 9660
attaaagatg gccaaatgtc tgagtcctga cagccacaat atggagatct agacttttta 9720
cagaccacag ggcacagggg cctcactaac agagttcccg gaagtgatga gtgtgctggg 9780
ggcttcctgg ttgaagagac actagaatgg acgagctggg agctaatttt ttgggctgga 9840
gtgtgatggc ctgcacatca ctgcctctgt ccctccattg tcacagctgc cccttaggag 9900
ccagctgagg caatttgtgg tcagagtgac tttgcacagt tgtcctgcct gtgttcagga 9960
agggagtttc tgtggtccct ttgaaaccac agaagagccc ctcgtatagc tctcaatgga 10020
gggggcaaaa cattcaaata actcaggaga taacacaact atttgttttt aactgtgagt 10080
ttttaggcaa tcacaaagat ccagatgtat gtccaagcct ctctttgcaa ttctatttaa 10140
cctcaatgtt gcaaccatag acctacctta cagagttcaa aaaaatatgc aaaaaccctg 10200
cctttcttct tcctcatacc ccaaaatgcc attctgaaca tttcctgtta gttaaaaaaa 10260
gatttccatg gtgttaccag gcactgtaca cagtctgtgt cccaagacaa ggaggtacag 10320
ttccacatgc gcccatgact gggttgggct ctgcactctc tctatacttt gagagcctga 10380
ttttctgtga ttgggcagag ctggcccacc tggtgcaatg tcctcctctg cctttcaaac 10440
atgttttagt catcaagatc ttcaaatttg taacccttc cagcttgatc cagcagaatg 10500
cagatttgga aaaacagaac gagtttaaaa tacatgattc taagaaacct ggaccagaac 10560
tatcaaaact tggtttccca gagaatatag caaatgggct cattggccaa tactatgaca 10620
ttggctttg agaaaagaaa ggctttattg caaggctggc cagcaaggag acaggagttg 10680
ggctcaaatc tgtctcccca gtttggggct tagggcaagt tttaattaca cagacgcatt 10740
tcttatgagt agcaggcaga gagcctccaa cttcttctgc ctaggtacca gcagcttaga 10800
catgatgcaa acctgggaag cacatactgt atttggagaa agtgattggg aagaaatgtg 10860
agctgagggg aggggctcag tgcccctgag ctacacttag tgatggcaga ggaaggatgt 10920
cctcccgcag gaggctgttc cacatctgct ctggttgtag ggggagctgg caggcattag 10980
cagcggcctc tttcccccaa gagaggcagc ctcctccaag ttttggcgac attatggccc 11040
tgcaatcata agggtttgtg agcatagtgc taaggaggga aatggagctg ctgttactag 11100
ttccacccca acacacacac acacactcac aagaaacctc acaagcaccg tattggaaga 11160
ctttgccatc caacctggga tttgacaggc tctagaagca gaatcataga ctcatgaagt 11220
tcccccaaag caggaatctt ccttacagta acccccaacc acccccctcc accgcctcca 11280

ccggctgctt cttcctgaac actgcagtgt ttggaaaact cacaaacttc caagcttgcc 11340
tttcctattg ttgcatggat tgaaagcttg cgttgtgtga agaatggcgc ttcctgctgt 11400
gcttagtttt atctcatata atctttgcac catttaatcc ttgcactcac ccactcatgc 11460
aactgccttt gcagagactg gaggggccgc tgtaggctga cctttccttc actgtaccta 11520
ttttgttccc tgctttattc ccctgcaccc aggacactgc ctggcacaaa gacaggtctt 11580
tataagtgta tgcaagtgaa taaagatata tatattatta ttgttatttt tgagacagtt 11640
tcactctgtc acccaggctg gagtgcagta gcgcaatctc agctgactgc aacctctgcc 11700
tcccaggctc aagtgattct catgtctcag cctcctgagt agctaggact acaagcatgt 11760
gccaccacgc ccagctaatt tttgtatttt tagtaaggac agggtttcac catgttggcc 11820
aggttggcct ccaactcctg acctcaagtc atcctcctgc ctcgacctcc caaagtgctg 11880
ggattacagg catgaaacca gcctagaaat acatactatt atttattctt gttttacaga 11940
taagcaaagt gagtcatgga gaatttggtt gaaagtccca aggtcaggag tcgtgaagct 12000
gggattaaaa cctaatcatc tgactttaga gagtagacac ttgctccatg catattgcct 12060
ccaattcatt cattcaagca ctccctgctc aagaagttct ttcttatgtt gagctgaaat 12120
ctgcagccct atgcgtttta cccagcagtc ctggtgctgt tccctaaaat cacttagact 12180
gtgcctgctc tttctgtgtt tacagtgtca gctgtaatat ccccctcttc ggcctaacgt 12240
ttctgaagtc ccttgccact gggtctcctc tcctcttcct gtgttctttc taagaacacc 12300
tatacagata ggtgtcttct gtacagggaa gctgttcctg agatccgggc atcgactctg 12360
ttagaataat ctacgtatga gttatttttt tgagaactat gtgtcattgc tgactcatat 12420
taactctgtg gttaactaaa atctcaagat ctctttatgt ttgttgagaa acttatttaa 12480
cttctctggc cctccgtttc cttcactgag cagtggagtg attgataacc tccacctgtg 12540
gttgctgaag gtcttgcaca agatgatata gttaaagtag ctagcagtgc ccacgtacgg 12600
cggatgcctc acaacggttt gcagccatct ctctatctgt gtctttgtct ctctctcaca 12660
ctggttttgg cttactgtta gcagctagcc gagataagtg tgtttatggt ctttgcatgc 12720
attgtttctg tagcatactg gaggattaca agaggttggg gagtgagggg gcggtgagga 12780
gtagacaaag gcagccaact cttccaagtt tagcttagaa ggaaggagcg gtaaacccta 12840
gttgaatgtt ggactgaagc aggtttgttt ttgttttgtt taaaggatag ggaagatctg 12900
tgcgtgtttc caggataaag aaaaggagag aatatgatat taaagattct ggaagtggga 12960
gaaggagcaa tgaaatacag acttgaagtc agtggcatgg acagggtcaa gatcacagtt 13020
agaggatgca gccttagaga aaaggaaggg gctcggttct ctgagcaagg agggaaagaa 13080
gagaggcaga tgcagagaag tacggcacat cgtgctgctg gttgtagaaa taacctctga 13140
cttttaataa agtcatccct cggtatccct gggggattag ttctatgacc tccctcggat 13200
gccaaaattc gtggatgctc aagtccctga tataaaatgg catagtattt gcatttaacc 13260
tacacacatc ctccatatcc ttttttttttt ttttttttttt ttttttttttt ttttttgggg 13320
agatggagtc ttgctctgtc gccctggctg gagtacagtg gctcgatctt ggctcactgc 13380
aagctccgcc tcccgggttc atgccattct cctgcctcag cctacaggtg cctgccacca 13440
tgcccagcta attttttttt tgtatttttt agtagagaca gggtttcacc atgttagcca 13500
ggatggtctc gacacatcct ccatatactt taagtaacct ctagataatc tctagattac 13560
ttgttttgtc ttttttttttt tttttctttt tgagatggag tttcactctt gtcacccagg 13620
ctggagggca atggtgcaat ctcagttcac tgcaacctcc gcctcctggg ttcaagcaat 13680
tctcctgtct cagcctcctg gtagctagga ttacaggccc ctccccaccc ccccccccaa 13740
caactggcta atttttgtat ttttagtaga gatggggtgt caccacgttg gcctggctgg 13800
tcttgaactc ctgacctcag gtgatctacc cgcttcagcc tcccaaagtg atgggattat 13860
aggcatgagc cactgtgtgt ggcctagatt acttataata cctgatagaa tgtaaatgct 13920
atgtaaacag ttgttatact gtattgttaa aagacagtaa caagaaaaaa aatctgtaca 13980
tgttcagtcc agacaaatgg ttttctgttt ttttttttttt ttttttaata tttttggtca 14040
gtggttggtt gactccagga atgcagaacc cgcagatata gaaggttgat tatgcgttca 14100
gaggcaggga ataccatctt gggttccaga aagaaaatga tcagcatttt ctgtcatact 14160
ctggtaaaaa cagatctttt gaatggacag gtgtattaaa ccctgtggag ctggctgggc 14220
ctggcggctc acgcctgtaa tcccagcact ttgggaggct gaggcaggtg gatcacgagg 14280
tcaggagttc gagaccagcc tggccaatat ggtgaaaccc caactctact aaaaatacaa 14340
aaattagccg ggcgtgatga cgcatgcctg tagtcccagc tactcgggag gctgaggcag 14400
aagaatcgct tgaaccctgg aggtggaggt tgcagtgagc cgagatcacg ccactgcact 14460
ccagcctggg caacagagtg agactccgta tctaaaaaaa aaaaacaaaa acctgtggag 14520
ctgatgaaat cctgcaggga gcttcacggt gacagcaaga ggagaaacac atccccatat 14580
gccccgcaga gtttgaagtc ccggctgcac ctctccccag cagcaggttg actctggaaa 14640
gttgcagcgt tcttacctac agagtgggaa cagtactacc cattgcacag agtgggtgca 14700
aagctctgtg acggaataca tggcaagtgc ccaccacatt gcctgggatg aggtgggccc 14760
ttcctttacg taagagaacc ctacagatac actcaaagtg ggcacattcc tacagaagga 14820
gtgttatttg tgtagaaaag aaaaacatga aaggctttta ttcctataca caataaagca 14880
ccccttaat gtctttttga ggaggataat atgaaattga tgaaaaggaa ccctgtggtt 14940
ggatccctga caatcacatg tatccctttt ttcactctta aaaaaggagt aaaggaataa 15000
aatagaaggg gagaggggge agagagacct tcaccgcccc cccccacccc ccatcatcca 15060

atctatagtc aaaccctcca gactgtgtct ccttggcatc tctgacaccc ccaccgccac 15120
caccccagtc aattcctatc ttatcccccc atcctggatc tgattctgct aagttcctgc 15180
cacactaaag acagggtggc tttctgatga caacattcct ctgcttaaac ctgtcagtaa 15240
ttccttgttg ctctcagacg gaactaagtt ctgaatttct tcacacggct ctcagcaagg 15300
tcacagtcac cctgctaggc cccaggggca aatctcaatg gtcatcttct tgaagacctg 15360
gctcagttat ttctttctca ttgaggctca cgaccccacc ttcttgcatg cctcaaacgg 15420
ccccttacca tgctcttctt tcgcccatag ctcagcacac cgtatcattt taatttatgt 15480
attttgctta atgtggatga tctgtctcct cctctgctgt cctcaccaga gcatcagttc 15540
ctcaaaccaa ggctctttgt tttgttcttg gatgcaagct aaatgtctgg catgtggcaa 15600
atggtcatag atacatgtca ttgaaagaat gattcatcac ctccctcttt ggccttgtct 15660
gtggttctac caaatcccat tccctcccca gtgccctcca ttccccctcc ttggctgaac 15720
attctgaacc acagacagtt ctttaccctg aacctttgca tattttgttc tcttagctta 15780
gagcggcccc tctccctccg tctgcttggc taatttctac ttgttcttca gattttatct 15840
tagatgtcat tccctcaagg aatccttctg tgactcaaca tggaattaag ttgcctcctt 15900
tgaccctgaa agcaccatgt actcaatctc atcttggcat gactcacttt gctgtgtgga 15960
atgtctgctt tccttgtttg tctattcctt tagactgtaa gatcctagaa agtgggggcc 16020
gtgccttgct catgactgtg tttctaacac caaacacagt gttcagtaga gagcagctgc 16080
tgagtacgtt tctgctaaat gacagttgat ggaggacatt tagggttgct tggaggtcaa 16140
gtcaaggagg catttaacat tctagtaaaa caaggaagta acaggctcct gaacatgccc 16200
acaatgaacc agatgcaaac cttttccctt ggcaggattc tttgcccata aagtggagca 16260
cgaaagcagg acccagaatg ggaggagctt ccagaggacc ggaacacttg cctttgagcg 16320
ggtctacact gccaagtgag tcctaaccct gatgttgcta ataagtgggg gcatgggcag 16380
gggggcctcc ttctaggagt gatgaccacc cttaatacca catgtctgtc tgagccaagt 16440
ttctgagcgc cagggaggtg aggaaggttg gacttcacca gagaggcttt gtggacaccc 16500
tttatcatct tagtgagtgc tagtgtcaaa acaaagggag tggggatatg gggcacattg 16560
gtggagggag gtgtgatctc tgcagcttca gaaagatctg aaagagtcat ttggttagag 16620
aagttgacct atttcctgtg gggttagacc agggttgcta ctgtgaacac cagccatgac 16680
tcaccagtca ccttcagaag ccacaggcag gacatgctga cgacagcctt caactcaccc 16740
acccccttgct cccctgcggg tggaagtctg gaggtgacac cactgcattt tctaacacgg 16800
gggctccttg agcaactaga acaagaacag aaagaatggg gacattagca ggtgctttcc 16860
ccctctctca ttctttttctt tgaataaaaa ggttgtttga aaacacctga gcggctccta 16920
aagatgggtg caatctattc gggatgcaaa tccgaatgaa tgttattcaa atgctcctct 16980
cttctttatg cagagtgtat ttcaaggctc agccagtggc aggcatgctg gggactatgg 17040
actacggact aggggcctgt cacagaggaa ggcctcatgc tagagagcta agggaggagc 17100
tggccttcag ttccatccca ggagcaactt tgatgttccc agagatcctt ccaaaggggg 17160
agtcatggtc acccaagaaa aatgtattca gaatgccaag aatggtgcaa actcaggaca 17220
aagattcaca ctgcagggtt ggagtccctg ggcttgctgc tggcaccatg ggagggaggg 17280
tcccccttcag gggtaccgtt ggtttcctgt gaattaaact ggcttcaagg gatctcgact 17340
gaacaggcct atatcacact cactgatata ctctctcttc agtccttctc ctcatctagg 17400
tatttttaat tgtttcagtg aggtgtaggc atgaggggat tggagggggc atctcctcca 17460
ttgcagtttt tcattggctg ctttgctccc tcagctccga aatcgctggg ccactctcga 17520
acgcattagt acggtagtca caggttgatt gcctggcccc ttgccctctg tgggcatttt 17580
ccctttcaga cagcccctga gtactcacag tgctgctaca gtgggccacc tagatctccc 17640
tctttctcca tgctcccacg tgctctgggc tccactccct tctcccaagc acttctgtcc 17700
agggctattc cagcagtctg acctcaagga aatcctttgc taaactgatt atagagaggt 17760
ttctatttta acatttaggc cttccatgta ttaattctca gaatcaattt aagatgttta 17820
aaggtgtgat ttaagacatt ttaaaaccat ttggaggaga gtacagaaat tatgtcactt 17880
gctgtcagcc tctttgcacc atctgcagag aaagatacta gagtcccgcc ttggacacat 17940
ccacatgcaa gaggtgcaaa gaaggtgtct ttgatgaggc aaggtcaaaa cttctcccca 18000
gacgaaatcc aaagaaagca ttcctactat gctatatcag tttggaaaga aaaacttctg 18060
ccaggtgact gcattctcac tggtcacatt gtgttcctat ggactcctca gctcaaccaa 18120
tttggagaag ttatggtgca atttcaccat atctggttag aagttaagtt tccaatttgc 18180
tggcaatgaa gaagaaatgg agcaggccag gctgtgtagt ttctgccacg tgcccccggg 18240
agtgaacagc tctgtttgta agaagccatg gtgcttagac ctgggctcgc tagttgccag 18300
cctccaaatt gcagaagtgc cctttggttg gtggctatgc tgtgtcactt gggaaggtcg 18360
tttggaagtt ccacagtcgt tgtggggtgc cagagattaa aaagcgtaag aggagagtgg 18420
aaagtgattg ttgctgcttg ggcatcccca ccgtgtgggt gctgcagccc agctctcaaa 18480
acccatgggt ctgtacactc aacctccatg agagggaagg agaaggatga gggaggggag 18540
agatagccat ggaaaggtag gaactaagca ggcagggtgg agagttttct gtaagacaaa 18600
aactgtctgg acactgctgc ggttctgtta caaagaccac ttcctccctg ggccagcaac 18660
atatctgtgt gcctgtctgg gttgtaaaaa gggtcaaaga tcaatgcagc aggcagctac 18720
atgctggaaa agccagaggc agctggtctg tttgcctgtg ccaggaaacc actgggaatg 18780
gggttgtgtg ttattctagg agaaagtcgt cccagcagca gcttctccag gggcatccaa 18840

gagcactgaa aaaggttgca agatgaccca tgaggctgca ggaagaaaag aacatgcatt 18900
taatcttgct atctgaaaag taagacatga agctttcctc atttttaata tacacatgga 18960
cagtagtatg tgtatatagt ttatatgcaa atatacttgt tataaggttg catgctcaaa 19020
atttttggtt catggggtgt gggatcataa atgtttaggg accatggcta tcaaggaaaa 19080
acagcatgaa ggataaatga tactggtgga ttaaaaagac agatgcatgt atttttagca 19140
taaaacacaa ctgctgactg atacagatag ctcaagattc tggggcagct gctgaacaga 19200
tacactagcc agtgtggctc atcggctcag acttggcctt aattaatggg ctgtccctcc 19260
acccatctcc catgagggca gagctgagcc agggtttgag agctaaaagg aattggacct 19320
ggactctgtt cacgtgtata ttttaattct aattaattca ttcttttgaa agacagagtc 19380
acactctgtt gcctaggctg gagtgcagtg gcacgatctt ggctcactgc aacctcggcc 19440
tcccaggttc aagttattct cctgcttcag cctcctgagt agctgggatt ataggcacat 19500
gcccccatgc ctgactaatt tttgtatttt tagtagagac ggggtttcac catgtcaggc 19560
tggtcttgaa ctcctgacct caggttatcc acccgccttg gcccctcaaa gtgttggaat 19620
tacaggtgtg agccaccgtg cctggcctgt tcacatgtat aaaacacagt ttaatgtcct 19680
attcccagcc aatgagcatg gctagagcag ccttggtcaa agtttggttt ttggagaaaa 19740
atccttgtta gctgacctaa gattcctctt tgtgagtgta agtaagcaca ggttgcagag 19800
aggagaaggg tctctggaga ggtgtaattt tctaaatgga ttacaagttc atggactttt 19860
aacaggtgtt acaggggata acaagttctt tatagacaga cttttgagga cgtttaaggg 19920
tattctgatt cttggttttc taagagggga atgtattatt taactacaga caccctacc 19980
gcccactttt tgcagagtgt atcaaaacat gtttttggaa taccaccctc atgtcgcttc 20040
tccctgcatc tcttatctct tggtgtccat tctagactca ctttctttct gtttttttatt 20100
tttatttttt tttgagatgg agcttcactc tgtcaccagg ctggagtgca gtggtgcaat 20160
cttggctgac tgcaacctct gccttccggg cttaagcaat ttttgtgcct cagcctcctg 20220
agtagctggg attacagcat gcaccaccat gtccggctaa tttttgtatc tttagtagag 20280
acagggtttc actatgctgg ccagcctggt ctcaaactcc ttacctcagg tgatctgccc 20340
gcctcggcct cccagagtgc tcagattaca gacgtgagcc actggtgcct ggcctagact 20400
cactttcaag tggcatagac ttgtaaaatt atttaaaggt gataggtcta caatgatcct 20460
gtcaattagt attgacacta ttattaataa actgttatta attatattta cttactttaa 20520
attaatccaa actaattaac ggaacactaa agagtttcta tgttttattc ccagaggtgg 20580
agaaaaatga aagggaatat agcaacgaat tcttttctcc ataaaaacat gaatagtgca 20640
gcacatcaag ttgaacatac cacagcaaat tgttgcaaga tctgctgagt agctcctatt 20700
tagacctcaa ggaatgagac tcaaaatggg ttcatcagtt ctgttttgca gaaaaaatag 20760
cgcaaaattt ctcaaaagaa aatccagaat aataataatt tgtcaatagg aaagacattt 20820
ccactggggg ttaagaagga agacattgga acaatgatag ccaccactta ttgaatgctt 20880
actgtgagcc aggtggcact tcaccttgtt tcattctcac aacagtctag ggaagtaatt 20940
actaatgtct ccatccacct cttgtagatg agcaaattga ggctcattga ggctaggaaa 21000
tgcacccaca ctcacatagc ccataagagg cagccatggc attgggccca gaccatgtga 21060
acttcaaaga ctacacgagc agccactggg cagctgtcat ggctaaagcc acttgaattc 21120
agcccagcag caacccccctc tccaggaggg gcacataagc ttgcagcttt gggtagaagc 21180
tgcacttgaa gtcctggatg gcgagaggga ctggcttgag ccagagccag gaacaaggct 21240
ctgagaatat tctggaaatc cacaggagga acccatttтc ttacagctgg gagaatttca 21300
ttcaactcca ggctgaccat gttttattag gaacgaaggt gacttgaact aatagtcagg 21360
aatggttgaa tacggaccca atgtcaaatc actaggcagt tcacatttct aatgagcaaa 21420
tcccttagac aattaagaat ttttttcctt ttgcataacc cagacaaaat cgctacttaa 21480
aaacaaacca aagacccgaa acatgagaaa gagaaggaag caggggaaat ctttggtact 21540
aataagtttt taaacaataa gagcaccaga tattttaccc catcagacac agaatgttat 21600
tcgaataacc aaaaaaggaa ttttttctct aagtttcttg aactggaaaa tgaatcatat 21660
tttctcagtc ctgaggctgc aattttgtgc ctctagtaac atataagaat agatgtgatg 21720
ccagtgccca gtagctgctg caattgttac ttggggacct gtttattcac taagcacttc 21780
accccagtga taaatttgta ggggcctcct gcccttgga gctcctaccg tgtccattag 21840
atcagtggaa attctgggat tcagagcact ttgcaaggtc agcaggggtc tgctctttct 21900
gtcctgttcc tggtttttgg ttgtgcctgg attccagggt aggtttctca tctgttacct 21960
tcatagactt ctccagaaaa ggatcttttg accatcagag gaccacgaag attccattgg 22020
tgaggcgcag ataacctgat ctctctgggt tctctgcagg gcacagatga agggctggcc 22080
attcccaagt tctcagtggt accactgagg catgagaccc taatggtttg catgagcagt 22140
ttgaaaattg catctttgtt tttacctata taatcacatg aaacccgtgg ttctcaaacg 22200
tcagcaggca tcagcatcac atggagggct tgttaaaaca gatttctggg ccccaacaca 22260
gagttttaaa ttctgaaggc ctgaggtggg tgtgaacatt tgcatttcta acatgttctc 22320
gatgctgctg ccgcctctgg tcccgagagc atgcctggag aactgccacc ttcgaccatg 22380
gactgtgaga attcacatgg acctcagaat tataatcagt ctctcagttt tacagataag 22440
gaaactaaat ccagagagat tgttttgcca atggtgaaca gctggttaaa gtcaggatgg 22500
agactttaat cctagtcaag tgacctttcc tctgtattta tttccctccc tttttatgcc 22560
tctcaagtct agttacactg tttttcatgg atgggcatat ttattgtcct gatctggact 22620

gcagacttct caggaggaca cctatgattt aatttagtat agttgaagag ttaacagaca 22680
tggctttgga gacagactga ttatggtgtg aatcccggct ttgccactcc ctagctggat 22740
gaccctgagc aagttattca gcttctccaa gcctgagttc cttattggaa acatgagagc 22800
aattgtgata ggcagaataa tggccccctc accaatcatg cccacatcct aatcctagga 22860
acctgtgaat atgttatgtt acatggcaag gggaaattca ggcagctagc cagttggcct 22920
taaaataaag agattatcct ggatgatctg ggtaggacct gatgtaacca caagggtctt 22980
tttaatgtgg aagaaggagg cataagagta gatgtcagag tcattcaaaa taagaaagat 23040
ttgatgggcc atccctgact ttcaggttgg aaggaggttc tgagtcaagg aatacaggtg 23100
acctctagaa gctggagaag gcaaggaaat ggtttctccc ctagaagttc cagaaggatt 23160
gcagccctgc taatatcttg actttatagc cctttgagat ttattttgga tttctgacat 23220
cctgaaccat agtaaaaggg tgtttttgt ttttttgaga cagagtcttg ctctgttgcc 23280
tgggctggag tgcagtggtg tgatcttggc tcgctgcaac ctccgcctcc caggttcaag 23340
tgattctcct gcctcagcct cctgagtagc tgggattaca ggtgcttgcc accacacctg 23400
gctatttttt gtgtttttag tagagacagg gtttcaccat gttggccagg ctggtcttga 23460
actcctgacc ttgtgatctg cctgcctcag cctcccaaat tgctgggatt acaaggcgtg 23520
ttgttttaag ccactcagtt tgtggccact tgttacagca gcaagaggaa actcatacag 23580
ttatcatgtg aactcacagg aatatggtga gttaaaaaga gaggaagggt gcaaaacatc 23640
cacggtagag tgagaactct ccagggagtg aggactgtgc ccagcataca gtgatcaccc 23700
tcttagtaag ctaagtttct gagcaccagc ttttttgagt tgactttgtt gtctttaaca 23760
tttgaagatc acccttcttt gctcagcctg gcttgcagac ctgggctgat ttgtggatct 23820
gatagaaaag tttccttagt tgggctcttc tccccgacca ccccccatgcc agtgtggcca 23880
catcctctgt ctgcattgct cactcttcaa ttccaagaag cgcaggggca ccgccaggaa 23940
caggaaccct gccagaggaa tacatcaaga aaccaagtct cccttacgca tcaccgtagg 24000
aacagagtta atggattatg aacatgtgtt tgctttatac cattgtttgt ttcccaggtg 24060
gcagctggct gccccatctt attgggtaga tgtaagtgga attacgaatg ggatttatgt 24120
ttcatgcacg atggtgatta ttaacttcaa ctttcaggta attttcagac cacattgcac 24180
taacttggtc tctgattgtt tttctccttg tttgtttatt ctgcagccag aactgtgtag 24240
atgcgtaccc cactttcctc gctgtgctct ggtctgcggg gctactttgc agccaaggta 24300
actcagactt ccctttgttc attctccttc tataaagtgc atctcaagga ggttcaaagg 24360
gcaggctttt tgttgaaagg actttgcctg acctctggct cccatctgtg aagccctgga 24420
gaggtgagag ccctcgggag gccgtgtttc aggcatgctc tgcacccgtg cagagcgcgt 24480
gtgataatgc attgctaatg cttgctccct ggtggctggc tgagagctgc tgtgctgaca 24540
agggtggttt aaggctaaat gtgactcaga atccttaagc agtgttagtt cagatacaag 24600
ggcattataa atgagagtgc ctgagggatc tattttggga ccgctgtcac ttggctcttc 24660
tgctaataag cttccagtgt ggtggccctc cttcaggcat gtttccactg agccacgggc 24720
tggatgccac atccccggcc ttcccacagt tatcagcagc ccacaggctt gacttgagca 24780
agttggaaag acaaatcaac ttccagagtt gatttaacat tgagtggaaa tcagtcatac 24840
ttttggtccc ctttcggggc cacgcctggc actgtgcctg gtggcagatc ggcatgaact 24900
ggccagcttc tgtggccctg gagggcacag gcagaaaggc cacgctcagt cccatgatga 24960
actgtttaag acttattgtt gtctccccgc tctgtaaagt agatagagtg gattttatgt 25020
cccttattac ctttcaggat actttgactc agggagataa agtaacttgg gtacagctac 25080
tcagctggtg aagaacacag gcagaatgag tgcctgggtc ttttgactta aaattctgga 25140
tttttcacaa agatcctctt actttattca tttacataat aaatatatat tgaagagcta 25200
ctctgtgcca agccctgtgc ctagatatac agtgataaat aaagagtagc ttctagaggt 25260
cacctggcgg tgaggcacag gccagctggc aagatggacc acagaagtca gtgaatgaag 25320
acaatgacaa gggtgggaag cgccatatgg gaagagaacc aagttcagtg atagagagca 25380
gaggtgaggc ggcagcagaa accacttaag ggacaccacg tggcactcct tctgtgctga 25440
gaaggctgtc agtaagctca ccatttattt cctatttttct ctcctgagtt aaataggaaa 25500
catgtctcgc attacttgaa aaatcaagtc aaactatgct cttactagga gttatggtcc 25560
tttttatgtc ttagatgatg cttgatctag atgaatgcgg acttgctgta gctagataaa 25620
tacaatggga gtttgaaggt gtttcgtagc cctggaaata tgtatttcct gtcaaaacaa 25680
gctttgtcat tgccagcaga caaaagcatc agtaaccttg gttgataatc gtcatttctt 25740
aggaataaag tagactgtag aatttttttt agcagaaagg aaacccaaag ataattctag 25800
tgcaaatccc tcactttata gagcagaagc tcaagtccca gaggaacaag tggcttgaac 25860
gaacatcaga attttagggg ctggatttgt accctcctgg tgccagcagc ccacttccct 25920
gcaggaggca ctcaccttcc ttgcacaggg gtatgagtgt ggccattttc cacccataat 25980
ctctgttagc tcatgttcaa ttgggttccc attgaaagaa aaatggacca gtaagttgga 26040
gcagaatcat tcagatggta taacataagg aaaaactttg cccaaggcaa atcgtgattg 26100
tgacagcttt gtgattttta gagaatagca tgggccaggc acagtggctc atgcctgtaa 26160
tcccagcact ttgggaggcc gaggcaggca ggtcacttga ggttgggagt tcgacaacag 26220
cctgaccaac atggagaaac cctgtctcta ctaaaaatac aaaattagct gggcgtggtg 26280
gtgcatgcct gtaatgccag ctactcggga ggctgaggca ggagaatcac ttaaacctgg 26340
gaggcggagg ttgcggtgaa ccaagatagc accattgcac tccagcctgg gcaacaagag 26400

tgaaactccg tctcaaaaag agagagaaag ctgaagttca cagtttctct tttgctttga 26460
ttttcttatc tgccggataa caatagtatt ttggaaggca ggaggaattg tggaaagaaa 26520
tgggttttgg ggagtggctg attggaggca aatccaagga cactcattgc tggtgtgtga 26580
ctccaggcag ttactcagct tttccaagcc tcagtttcct tattgtaaaa caggaccatg 26640
gtctagctag tagcattcct atggtgagtg aaataatatg tataaagctc ctgacacagt 26700
gcttggcata tatcagattg agccatgtaa aactgccaat atctggctat ttatgaccta 26760
caaaaatagc atttcatatg attccaccta acatctgaag cgcaataaat gttattattg 26820
ataatgcagg tggtggtgat aaagttttga aatcagaaag acctggcttc aaattccacg 26880
ccttcactgg cctgacttat tttcattcat ttgacaaata ttattttgaa caccctatg 26940
tgccaggcac tatgccaggc tcagagatga tctaggaaaa agacagatgt cctcatctgt 27000
cttaggctct tgtggcctaa gcctaaattt cctcgtctgt caaatggtga cagtaacaca 27060
ctccttacca gagagctggg aggattggag actcaagttc ccaaaacgcc aggagcactg 27120
cggcaggtga aaagtattcc ctcaatggcg gaagtgttta aattgctttt atatctgtag 27180
ctctagataa cactagttcc agcttagtta actcccagct ccaagccttc aggacttcat 27240
agagttattg gggtgctgct cttggcagtt tcccaaaaag ctagaatgca gagggaatct 27300
ccttcccaaa aagctagaat gcagagggaa tctccttccc aaaaggctag aacgcagagg 27360
gaatctcctt cccaaaaggc tagaatgcag agggaatctc cttcccaaaa ggctagaatg 27420
cagagggaat gtccttctct tctaaatggt agctgttagt tcaagaaagg ttaaacattg 27480
tgctgtgggg aggctcaggg gtgaagggtg tacttttaag agaaccagtt tcagagctgg 27540
gtttggggtt taagccctac cctctgcccc cttttacgag ctgacagcct tatgcaagcc 27600
tggttgacca cctgaaccca cgtttccaca tctggaaata gaaatgtggg tactagttat 27660
gttgaaagga ctcaggttag atgatagata tgcaaatacc ttggaaacca ggagtgtcca 27720
gtcttttggg ttccctgagc cacactggaa gaagagttgt cttgggccac acatagaata 27780
cactaaccct atcaatagct gatgagctaa agaaaaaacg ttgcaaaaaa aatctcatat 27840
ttttaagaaa gtttatgaat ttgtgttggg ctgtattcaa agccatcctg ggccacgtgc 27900
gacccgcagg ctccgggttg gacaagtttg ttgtaaacaa tgccatgatg ccggcataag 27960
gtcgttacca gtattaggaa ggttctcagg tttcctctag cccttgggct cttttcctga 28020
agtgcgtgtg tcttctgcta gattttgtga ccaatgttga ttgcctaatt gggctaacag 28080
catgttttgg tggctacgaa actgacacag gtgttttcat ttctccactt agttcctgct 28140
gcgtttgctg gactgatgta cttgtttgtg aggcaaaagt actttgtcgg ttacctagga 28200
gagagaacgc agaggtaggt aactgggact actaaagaac tgtggagcga ttcctgattt 28260
ttgagcagga agagtgacaa ttcaaaacag tatttgacta gattcacggc tccgtagcat 28320
ccccttgggt gggagcggga aggctgacta ggacctctga ttcttctctc cctgagcttt 28380
gaaggctctg aaaatacagc tgggggggact tgcccagttt tcttattaag caattcctcc 28440
gcatggtgct ggctttcaaa gggtgcttca gtgctgtttg ctgcacgtgc cttgcagccc 28500
cacaccctgc actcccgccc tgcagagtct ggcgctggaa tgacatttta ggtctgggtt 28560
cccaggcctc ctgagagtga aatgtttcat tgtttgtcta gagaaatgag aactaaagct 28620
tgcaccttgt gataagttgt cctgaggaac atatctttca gggaccagaa gaaagaatgt 28680
tgggaaaata agatgcagta agatgcagac atgacagcag ggtgcagcgg ctcacgccta 28740
taatcccagc actttgggag gctgaggtgg gtggatcacc tgaggtcagg agtttgagac 28800
cagcctggcc aacatggtga aaccccgtct ctactaaaaa atatacaaaa cattagccag 28860
gcatggtggt gggcgcctgt aatcccagct actccatagg ctgaggctgg agaatcgctt 28920
gaacccagga ggcagaggtt gcagtgagcc gagattgcgc cactgcactc cagcctgggc 28980
aacaaaagca aaactccatc tcaaaaaaaa aaaaaaaaaa aaaagattca gacacgagac 29040
tgtgaaactg actagcatca ccattgcatt gtttatagat gttgccagac agaaagcccc 29100
aaagcagcac agtaccttcc tgacatctgg actaggaaat ctagatttta gtaaaataca 29160
tgctaatact tacagaagaa atgtcggcgt tagagtatgc cgtcagttcc ttagagattg 29220
caattcctaa tgcactagta tggtttcagg tgccaggaac acgttctgtg aggctgctgc 29280
cccaggtgct gaccccagcc ttccacacca ttttccttcc ttgtgttcac agccgctctg 29340
tcttttacaa tagcacccct ctctagtggc taatgggctc tatgattaga tagcatcctt 29400
cagtagtgat aaaggcagtg acatcctagg gaggtcagcg ggtgaaagcg ctatatctgg 29460
aaaacctgag agcctgtgaa gctcaaggac ttgacggggt tagaccgtga gccgggctgc 29520
agctggaaaa agaatgactg ttctttcagc agatccttcc ctgtgccatc tctttcttca 29580
ttcctctcta gtggcattct tatttatcct ctaaaaccac aattccatta tctctcctat 29640
tcttatcaac actgccctaa atgatattct ttattctctt ttgccctgga aaacctctat 29700
catgcctttt cccatgtgat tacctcgtta agagtggggg tggaatgtct agcaatgaaa 29760
taagagggtc ttctcttttg cctggctccc tatgcagccc tatcttaccc cctgcaaagt 29820
cccagggatg tggctcagtc actgctcctc tcttcatctg tcaccacttg cttgagatcc 29880
tacagctgct ttaattccga gaccatctgc agaacatgac aaaatttgtc cacctaccca 29940
catgtccttt taactttaaa ggctttacta actgattcct attagggaat gaacagaggt 30000
ggcaaaaata aacaatagga gattgattta caagaaatct ttaaaatagt agatttcttc 30060
ggacctcatt gaaatataaa tggcctgcct tcttgtgtcc ctccctggtc tccctcttta 30120
ggtgataaga agaagatcct gccagcccca taacccgcca tctgcgcggg ttctagaccc 30180

ccttctcctc ccctctggcc gtggtaggca ttactgatga atcatggtgc tctttcttcc 30240
agagaccaaa cctggcctcg gaatccttct taacacagat actgcttaac acaaccactc 30300
tgagcagctg tcataagtag aagtaataga tactagaaga aatgtctaag cctaatctag 30360
accaaaatac ggcctgatat agatgcaagc cagaggggct ttatggttaa atgcaaggag 30420
attttcaacc ctgccgtcta gaagctactt gctgagatct tcttcagttg ggcccatctc 30480
ctccccaggc ctctcttctg ttcctgggct atgtcacact tggactctgc agacacctaa 30540
tgctcttggg acctgcttta gttcttgacc tcaccaaccg aggaggaatt gctcgatgag 30600
atccttcccc cggaatttct ctcttgaacc ccagatggtc cgttgccccc ttccagaagt 30660
tgctccagcc ctgtccgctt aggaagttca gtgtcatcct tgatccagtg ggtagggaag 30720
acattccata atgtatgccc cagtctgagc ttcttccttc aggcttcagg ctgccctgcg 30780
aggattttgc agctcccttt ttaatgccct ctagaagttt ctggctctta ttttcagccc 30840
ttcatcctac tctctctgac cccttcctct atcctgttta gttcacctgt agcagttact 30900
acccagcagt gaaggatgaa tcttggtttc gtttcttttc tcttcttttc ttttttctct 30960
tctcttttcc ccttcccttc ccttccctcc cttcacatca cctcatctca cctcacctta 31020
catagtcttg ctctgtcacc caaactggag tgcagtggcc tgatcttggc tcactgcaac 31080
ctccacctct tcccaggttc aagtgattct tatacctcag cctcttgagt agctgagact 31140
acaggtgtgc actaccacac ccagctaatt ttttgtattt ttagtagaga tagggtttag 31200
ctatgttggc caggctggtc tcgaactgct gaactcaagc aatctgccat ccccggcctc 31260
ccaaagtact gggagtatag gcataagcca cccatgatgc ccagcctgaa tcttggtttc 31320
ttccccattc atttaagcta ttacctgggc ctgaactcaa tggcacctgg caccaactgg 31380
caactgactc ttggtctttt attacctacc ttccctagca ggcactgggt tgctccctct 31440
tcctatccca tggagtcctg tcctctgttg gggctcctac tgatcctctt ggcaatatga 31500
agttctcagc tcaatggtgg gtgggcaatg actgccaact cttgaggcca atgaactcag 31560
gtttccccac tcctcctcct cctgagttgc tcactcactc ctcattcact caacattgat 31620
tcagtagata tttgctacct gctctgtgcc aggtaccagg tcagttgctg aaggagtaac 31680
agtgaacatg acggagtctt tgtccccaag gagacccaag gtgtctccta gagccagggg 31740
cacattgcaa gaccaaatat attcaactta ccaaaataat catagaccta gttctcaaaa 31800
agcaagaaga ctgattcctc gttgtcattt ctcctcctca gcatcaatgt tttagagtct 31860
gtgggccccc ccaagtgtgg agtatggtgt tacttcacca gagtttgagg agaaacattc 31920
ttcttttgga aggccgggga gcatagatgg atatcaaggc tgctgtttct aaaagcgaaa 31980
cccaccaaac aacagtatta gaatcatctg tggtgcttat taaagataca gattcctggg 32040
ccccatccca gacttatgaa tcagaatctc tgccagagga agcctgagaa tttgcattct 32100
cagatgattc tgcattctca gataacacat tctttaggtg attcttacac acactggagt 32160
ttgggaatcg ctgaaggctg ttcacttctc ttttctgaga aatgattcat tcatttcaga 32220
aatatttgca gaggtcctta tttattggag atttgtgggt gggcagagga gaaatatctt 32280
gtcctcacag agcttacaat ttttatttcc tttagaggtc accaggctta aaatgacact 32340
tccctaaatt ctgaaaagaa cagattttta aaacaagaag ggactgtaat gttttctgtt 32400
cctacctcgt attttgttca cattaagaac ctggggtggg aagtggagga ggggggggtga 32460
ctggcggggg gccacagaga gctgagctgg ggtggtctcg aactcctgaa ctcaagcaat 32520
ctgccagcct cagtctccca aagtgctggg attataggca tgagccaccc acgatgcctg 32580
ggtggaactc agggctctgg atgcctgggc gcccccatct cccacactac ggcgcctcat 32640
cctagaagtg gttagcacct ttgagatggg aattatttag caggatgctt ttgtgttttc 32700
atgtaagttt tatgctgcct gtggagggca cagctgtttc aaaaataata accaaatcct 32760
ggtctccgaa gtctgaaggc atcctttgcc ctgcagtgca aagcacggga ttctggcctc 32820
acacaggcag gtctgaactc ctgtgttgcc tcttgctggc tgtgggacct gaggcaaatc 32880
atgcaacctc tcttttctgt ttgcctagat ggaaaatagg tttacaatac gcccccatag 32940
gatggctgtg agaattaaag gaagtcatgg gtgtacaata cctggccccg aaagatgctt 33000
aataatttaa ttctgacctt cctcactcat ttaggattat gtaccaactt ttagaaacaa 33060
tgaaagatta gtgagtcttc tgtggttggt ataaaaaaaa aatagaaaca tgaaagagat 33120
gtcctccttg ttcaagggct aatgaccctg gtgtgcgctg tctaggcccc caaggtcttc 33180
cttccctgct cacagcattt caggttctcc gcagctttgc tgagcctggg tcaggttcgg 33240
tatctgccca ccatgctcac ctgccacagc tgtggcccca tttccaaact tcagagactt 33300
aaaggtgcag ctaatgatgt gcccggcctg gggtcacatt ccctgagccc tgcagacaag 33360
ggagcaggag gctgagctct tatcttccac accctgtgca cagcctggga agagttaaag 33420
caccctagtc ctatgctgcg agggccacat gccctgagac cttggaaaaa atcctacctg 33480
aattgaagag catcactatt tcatcaggag gcgctgccat ttcatttttc acttcggttt 33540
tatcttgagt gtaaaacagc ttcgcaaatc actttttctt gtttctgtaa tgagcatatg 33600
gtggcctcat tcgtgtgata aatctgagcc accacgatat ttgacttttc acaatttaat 33660
ttatctgaac cctctattct ctggctaaaa aatatccctt acttggactt ctttatttta 33720
ttttcaattc ccttaccagc actagcaggg gactctgtac tcatctgctg gcgctgccat 33780
aacaaagcac tgcagcctgg ggggctcaaa ccacagaatt tattctctca cagtcctaga 33840
ggctagaagt ccaagatcaa agtgtgggca gggtcggttt ctcctgcagc ctctctcctt 33900
ggcttataga gtgccacctt ctacctgtgt cttcacatca tcacctcact gagcatgtct 33960

gtgtccaaat ctccccttct tataagaccc cagtcatact ggatgaggat ccacccatat 34020
gagttcattt taccttaatt atctctttaa acaccctgtc tccaaataca gtcccattct 34080
gaggaactga gagtaaagat tcaacatatg aattttggaa gggacctaat tcagcccaca 34140
acaccctctt ttgggatgtt tattttcccc cttaaggagc tagttaggat gtcttatctc 34200
atgaacatga ctgtgaacag gaaaacaggg agagaatgaa gctggccaag gaacagggct 34260
ggtgtcagct agcagtgctt ttctgatgtg agtgggtccc acagggagct tgttaaaatg 34320
cagattctga ttcattaggt tccagaggga cctgagattt cccatttctg acaagtttcc 34380
agtgtggggg ctgatgctgc tggtccacgg accatacttt gagtagcaag gagcttgata 34440
cataatggct gagtgacttt cagactcctg ctgtagaaaa attatgagtt ggctgggcgt 34500
ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggtgggcaga tcacctgagg 34560
tcaggagttc gagaccagcc tggccaacat ggtgaaacac catctctacc aaaaatacaa 34620
aaattagcca ggtgtggtgg caggtgcctg taatcccagc tactcaggag gctgaggcag 34680
gagaatcgct tgaacccggg aggcagaggt tgcagtgatc tgagatcgtg ccactgcact 34740
ccagctgggc aatagagctt gactcagtct caaaaaaaaa aaaagaaaag aaaaagaaaa 34800
attatgagtt atattatcag catatggggt gcctttcaaa ttgataaaat ttctaatatt 34860
aaacctgtgg atgccaaatg ctgctctctg attatggcag gaaacggcac ttggcagtac 34920
gaagttagct gttgggctga gctggctcat cttgttgtgc ggtcctgatt gcctaaagat 34980
gccttcccag gatctttact aacaatcctc ctgagtcatt tggactttcc caacctgtta 35040
tcacctctca gatgggccag ccatggaggc agtcagagga gggctctgca gagggagggc 35100
agaaacaggg tggcctctgc atgccattag gaggtcacat ctcactgggg gatgcagttt 35160
aggatttagt gccttggaga gaaggataga gtgtattaaa acatgtctcc gctaggcatg 35220
gtggtttacg cctataatcc cagcactttg ggaggccgag gtgagtggat tgcctgagct 35280
caggagttca agaccagcct ggctaacatg acgaaacctc atctctacta aaatacaaaa 35340
agttagctgg gagtggtggc gtgcgcctgt agttgcagct acttgggagg ctgaggcatg 35400
agaatcactt aagcccagaa gactgaggtt gcagtgagcc gagattgcac cactgcactc 35460
cagcttgggc tacagagtga gactctatct caaaaacaaa gaaacaaaca acaacaataa 35520
caacaaaaac caagtctctc cctccactca aaaatgcaag ggcctgtctc ccattgctgg 35580
gtgcccaggt ctcatgaatg tagacatgaa ttattccagt cagcctcagg agaatagaat 35640
gagccctcag atgccgaagc acctttcaga ttccaccggt tttatcggct catttaaact 35700
tcacttctaa cacagtcctg cattacacac gtgtctgtcg ttatgggcag ctgcagagag 35760
ggtcttaatg gtcctaatgc tcagtgagga tgcccaatgg tcaacagaac ctgccatctt 35820
caggccatca aggagctctg gagttaagga aatcatgaga gcacagaggg gcgggtacag 35880
cagagccctc gtggtaatgg gttttgaggt ctaggctctc ttcgcttggg tttgaaataa 35940
gttcaatgac tagtaatagc tgagacactt ctacccttca aatgaagtaa atgggaaaat 36000
ggagcattgt tgagtccagg gagctataat ttaaacccca tatatctaaa aggggtaaca 36060
tttttgtgtg tgtgaaattg gtgtcattcg cactgcatct acagttttct ttttccttct 36120
cttccagcac ccctggctac atatttggga aacgcatcat actcttcctg ttcctcatgt 36180
ccgttgctgg catattcaac tattacctca tcttcttttt cggaagtgac tttgaaaact 36240
acataaagac gatctccacc accatctccc ctctacttct cattccctaa ctctctgctg 36300
aatatggggt tggtgttctc atctaatcaa tacctacaag tcatcataat tcagctcttg 36360
agagcattct gctcttcttt agatggctgt aaatctattg gccatctggg cttcacagct 36420
tgagttaacc ttgcttttcc gggaacaaaa tgatgtcatg tcagctccgc cccttgaaca 36480
tgaccgtggc cccaaatttg ctattcccat gcattttgtt tgtttcttca cttatcctgt 36540
tctctgaaga tgttttgtga ccaggtttgt gttttcttaa aataaaatgc agagacatgt 36600
tttaagctga tagttgaggg gttttgttaa tggcttttgg gggatttatc tctataccca 36660
caaacgacta gtttgttttc ctcaaactaa atgataatat taaaaataca catcctggcc 36720
aggtgtggtg gctcatacct gtaatcccag cactttggga ggccgaggca ggtggatcac 36780
ttgaggtcag gaattaagac cagcctggcc aatatggtga aagcctgtct gtactaaaaa 36840
tacaaaaatt agccaggtat gctggtggat gcttataatc ccagctactt gggaggttga 36900
ggcaggagaa ttgcttgaac ccgggaggta gaggttgcag tgagccaaga tcatgccact 36960
gcactccagc ttgggcaaca gagtgagact ccatctcaaa ttaaaaaaaa tacacatctg 37020
gcttctggaa aaattacttg aagatctttt atgacatcca tccctcttca cacagccatg 37080
tgaattaggt tggtatcttc atatactagc atcgtgccca gcacttccat gttatacagt 37140
ttaaaaggtt ctgtaattcc ctgtgggaac ctaagataat gcgaggaccg tcatacgtgc 37200
ccccaaatat tggcaaacca atgaataaat gaatgaatga gtttatgaat cgctaactgg 37260
ctgtatttaa tgaagtatgt gtgttgagcc atttcccaca gtgtggacag atttgtccca 37320
caatatgggc ctcttcccaa aggccctacc acctaatgcc atcacactgg ggatttgatt 37380
tcaacatgtg aatttgggga gagtgcaaac actcagacca tagcaccatc tcagtaaatg 37440
tcccactggt cactcagttc atagtgacag tgatccagcc actgtcatga caggtgccac 37500
ttggcagaaa cagcacagct tggaagatgg cggggtgtag tcaagattcc aggatcccca 37560
acagagaagc cagctcttat aggggagcca ttcatcagga ttgaactctc aatcgagctg 37620
gacagtaata ggtgggtctg tgttattccc caggtgagta tcatgacagt cacaatccta 37680
ggaaggatgt gaagcctccc ccagctctcc tccagttgcc tgcttgggca gcagagatga 37740

tggaatgtgg agtctggcgt ggtctgaggc ctgaatccat gtgcctcatg tatgatgctc 37800
aggcaagagg atctctcaat tcaagggaga gggcctgaat gagccttgct ttccaggcct 37860
gtctgatggt ccaggctgaa gcccctcctg gcttgcactg ccagacctca tccagcagga 37920
gctccttggc attgactgct tcaggatagt tgcttctgct ctgagtgctc tctaaagagc 37980
agtgctctac catccaagct gggcttttct tttcttcttg ctgataggga aggcatggga 38040
cattgcagga tggaagtggc ccccaggcct tctcatgcct gggcttggtt tggaaggtgg 38100
tcaggtgatc aataatcctg attggcctgg cattgaggag ttttcctggg atgtggtcct 38160
ttcggttttt taaaaattat ttttattgat acacatattt gtaggtattt gtggggtgca 38220
tgtgatactt tattatgtgt gtggattgtg taatgatgaa gtcagggcat ttagggtctt 38280
catcaccttg attatcattt ctatgtgttg agaacatttc aagttctcag ttccagctat 38340
ttttgaaata gacagtccat tttgttagct acagtcaccc aacccggctg tcagacattg 38400
gaacttactc ctattgaact gtgtatttgt acccattcac caaactctct ttgggctttc 38460
agttttacaa ctgggatgat cctgggaaaa ctaaagtaaa tcagacaccc gacgtgtgag 38520
ctaggttata atatgcccag tggaccctgg ggacatctta gctttcagag gtcatgctgt 38580
ccaagctgac tgtggggctt ccagaaggtg gggagaggaa atgatgcaat ggcccatcag 38640
aggcactact tggggcctgg ggccagagtg catgtctaag gcattaaggg gaggggagag 38700
cagccttcat aattatgaag aggagtctca ggtgcacagc ttctgatgag ggacagcttc 38760
taattgaaga cagcattgtg taatgctcaa actccctgtc ttcagagtgc ctgctgtatc 38820
ccaccatcag ttctgtgact tctccctaag cctcaatttt gcatgtgtta cattgggata 38880
ataatagtgc caaactcatg gggttgtgag gaataatgag gtaaagcaat tgaaaaggtt 38940
tagcacaata taagtgctca ataaaagcca ttattattat tttattacac tagttttcaa 39000
ttcctgcata gcaaattctt gcaaatgtag ggactcaaaa caatataaat ttattatctg 39060
acagtttttc tgggtcagag gtcttactag gctgtaatca gagggcaacc aaagctgtga 39120
tctcagctga agctcaggat tctcttccaa gctcactggt tgttggcaga attcagttct 39180
ttccagttgg aagactaaag cctacagtct tcagtctcta gaagcctttt ctctggcaca 39240
ggtttctcta caacatggcc atttatgtct ttaaggccaa taggagaaca tgattagcat 39300
atttttttta agtgaacttt agacccttttt ttaaaggcct atctgattag gccaggccca 39360
agtgagcttt aagtcaactg attagagatc ttaattacat ctgcaaagtc ccttcatgtt 39420
taccgtataa cataacttag tgaaaggagt gaaattgcaa ccaggttctg cctgcactcc 39480
acggaagggg attctgcaga agtgtgggtc acgggggggt tattttggga ttctgcctac 39540
gtcactgagt caaaagaagc tgaatggttg tgatgctgag gtttttgggc agcagcagtg 39600
tgtgtgtgtg agtgaattca tacgtatgac cacctgggaa gaaaggaggc tgtggtttcc 39660
tccacctcct ggcagacaga gaaatttctt tttttttttg agacagggtc tggctctgtt 39720
acccaggctg gagtgcagtg gcttgatctc tgctcactgg ctcactgcag cctctgcctc 39780
ccaggttcaa gtaattcttg tgcctcaact ccaagtagct gggattacag acacacactg 39840
ccacgcctgg ctaatttttg tatttttagt agagacgagg ttttgccatg ttggccaggc 39900
tggtcttgaa ctcctgacct caagtgatcc gcccacctca gcctcccaaa gtgctgggat 39960
tacagacgtg agccaccatt aaccatttttt ctatctcctg tgggaaaggg cacagtgaaa 40020
gaacagatga agctgagaca tacaagtgaa ctcctccctc ctctccattt agactaaaat 40080
aggattattc atactgagat tctccctggt tgcaaagaga taatctgtgc aactgggttt 40140
ttacaattat ccctacccta tgctttcctc atctgtcttc ctcgtagtca gctcaggctg 40200
ctataacaaa acaccataac tgggggcttt tgaacaacaa aactttactt ctcacagttc 40260
tagaggctgg aaatccaaga tcaagtttct ggcagattcg gtgtctaatg aggtcctgct 40320
ttccagttta tagacagtgc cttatcgcta ccgccttaca cagtggaagg agaggacgag 40380
aagctccttg ggctttttttt tgtttctttc tttctctctc tctctctttt ttttttttt 40440
aataaggtca ctatcttagt ccattttgtg ttgctaaaag gaacatctga ggttgagtaa 40500
tttattttat tttaaaaagt ggccaggcat ggaggcttat cctgtaaccc taatccttta 40560
ggaggccaaa acagcaggat tgtttgaggc caggagttca agaccagcct aggcaagata 40620
gtgagacccc atctacccca tctctactaa aattttaaaa aattagctgt gtgttgtaaa 40680
gtgtgcttgt agtcccggcc acttgagagg ctgaggtggg tggagttcaa ggctgcagtg 40740
agatatgatt gagccactgc actccaaccc gggtaacggg gcaagacctt gtctctattt 40800
aaaaaaaaaa aatctttatg tggctcacta ttctgggtgg ctggaaagtt caagattggg 40860
catctgcatc tggtgacagc ctcatgtcgc ttccagtcat gggggaagac gaaggagagc 40920
tggcacgtgc agatatcacg tgttgagggc agaagcgaga gagagagggg agagatgcca 40980
ggctcttttt aacaaccagc actggggaaa ctaatagagt gagagctcac tgactcctga 41040
gggaggacat taatctattg atgagcgacc tgcctccatg acccaaacac ctccaacgat 41100
accccacctc caacactgcc acactaggga ttaactttca acttgagatt tagaggggggg 41160
aaacttacaa actatcgcag gcactaatac cactcatgag ggctccacct tcatgaccta 41220
atcacttcct aaaggcctta cctcttaatc tcatcacatt gaggattcga tttcaacttg 41280
aattttgggg ggacaccaac attcaggcca tagcatcatc tcaataactg tcccattggt 41340
ggtcactcag gccccaaaca aaggaacctt cctccattcc tttccgccct cccacccaca 41400
gtcaatcatc cccaagctcc atcagctcca cctttaacgg ccaacccacc tctgccacat 41460
ctcaccatct ccactgctat ccctgtcacc tgggcccacc attctctctc ctggacagtc 41520

tccatagcca cctctgtcag atttatttta ttttttatt ttttttttg agacaggttc 41580
ctgctctgtt gcccagactg gagtgccatg gcatgatcac atctcactgc ggcctccatc 41640
acctgggctc aagcaatcct cccatctcag cctcccaagt agctgggact actggcacca 41700
ccatacctgg ctaatttttt gttgttgttg tttaattttt aatacagatg aagcctcact 41760
atgttgccca ggctgctctt gaactcctgg gctcaagtga tcctccggcc ttggcctccc 41820
aaagtgctgg gattacaggc atgagccacc gtgcccagcc catcagatgt taatgctaca 41880
cgcacttgct taaaatcccc cagataattc tcgctgctct tggaataatt cccacacacc 41940
ttggcgtggc catgcaggct ctgtgccatc ggatatgtcc ctgccccctc tcccaactcc 42000
tcctttcgct tgctcgttca ctcagttcca gccacattgc cctgggagct gctcccacca 42060
tggggcttcc taatgcactg gtctctctca tgcagtgggg cctctccctc cttttactca 42120
gtgtctccca gcacccacct cctccagagc cttccctgac caccacacct acacctaggc 42180
ccttcctcct ccacgctccc tcctccaccc cggcctccta cccacgtgtc acttctttat 42240
actcgctgcc acctgaaatt agatcattta tttacccctt tatttgttca gtttgccttg 42300
tccgttagaa tataagcttc caaagggcag gagctttgcc tatattgtta ggccgggcat 42360
acaatgagca ctcaaaaaaa tatttgatga gtgtatgaaa gaacagactg ggttatgtaa 42420
ttgtgcctac ttacctatat gaccatgtgg tggggtttat ggtgggtgtg gtggtgatgg 42480
ctatagggct ataagcaaat ttgggacagg gagtctaaga aatgttctta aattttagta 42540
agcaaagcat cctctacaga acctgtctta aaacatgaaa gttccttagt gctaccccca 42600
gaggtatgat ttggtaggtc aaggataggg cctggaaatt cacattcttg ttaagatgtt 42660
cttcatccgg ggtttgttga ccacctttc agaagatttt tgctctgtag ctgtactacc 42720
caatgcagta gttcgtagtc agtgtggctc ctgagccctt gaagtgtagc tcctctgaac 42780
tgagacgtgc tgtaaatgta aattgcacac cggagtttga agagttaata caaagaaaaa 42840
ggaatgcaaa acatctcatt aataatgctt tacactgatt acatattgaa atggtaatct 42900
tgtagatata gtgcgttaaa taaaatatac tgttaggctt aatttcacgt ctttatactt 42960
ttaatgtggc tactagaaaa atttaaataa catattcagc tcacattata ctcctattga 43020
acagagctga tctataagtt ccatggaaga tggcaagtct tcgcagctg 43069

<210> 2
<211> 875
<212> DNA
<213> homo sapiens

<220>
<221> 5'UTR
<222> 1..74

<220>
<221> misc_feature
<222> 75..77
<223> ATG

<220>
<221> misc_feature
<222> 558..560
<223> stop :TAA

<220>
<221> polyA_signal
<222> 851..856
<223> AATAAA

<220>
<221> 3'UTR
<222> 561..875

<220>
<221> misc_feature
<222> 74..584
<223> homology with sequence in ref embl : X52195

<220>
<221> misc_feature
<222> 354

<223> diverging nucleotide C in ref embl : X52195

<220>
<221> misc_feature
<222> 555
<223> diverging nucleotide T in ref embl : X52195

<220>
<221> allele
<222> 197
<223> 10-33-175 : polymorphic base C or T

<220>
<221> allele
<222> 453
<223> 10-36-164 : polymorphic base A or G

<220>
<221> allele
<222> 779
<223> 10-498-192 : polymorphic base A or G

<400> 2

```
acttcccctt cctgtacagg gcaggttgtg cagctggagg cagagcagtc ctctctgggg     60
agcctgaagc aaac atg gat caa gaa act gta ggc aat gtt gtc ctg ttg      110
              Met Asp Gln Glu Thr Val Gly Asn Val Val Leu Leu
              1               5                  10
gcc atc gtc acc ctc atc agc gtg gtc cag aat gga ttc ttt gcc cat      158
Ala Ile Val Thr Leu Ile Ser Val Val Gln Asn Gly Phe Phe Ala His
        15                  20                  25
aaa gtg gag cac gaa agc agg acc cag aat ggg agg agc ttc cag agg      206
Lys Val Glu His Glu Ser Arg Thr Gln Asn Gly Arg Ser Phe Gln Arg
    30                  35                  40
acc gga aca ctt gcc ttt gag cgg gtc tac act gcc aac cag aac tgt      254
Thr Gly Thr Leu Ala Phe Glu Arg Val Tyr Thr Ala Asn Gln Asn Cys
45                  50                  55                  60
gta gat gcg tac ccc act ttc ctc gct gtg ctc tgg tct gcg ggg cta      302
Val Asp Ala Tyr Pro Thr Phe Leu Ala Val Leu Trp Ser Ala Gly Leu
                65                  70                  75
ctt tgc agc caa gtt cct gct gcg ttt gct gga ctg atg tac ttg ttt      350
Leu Cys Ser Gln Val Pro Ala Ala Phe Ala Gly Leu Met Tyr Leu Phe
            80                  85                  90
gtg agg caa aag tac ttt gtc ggt tac cta gga gag aga acg cag agc      398
Val Arg Gln Lys Tyr Phe Val Gly Tyr Leu Gly Glu Arg Thr Gln Ser
        95                  100                 105
acc cct ggc tac ata ttt ggg aaa cgc atc ata ctc ttc ctg ttc ctc      446
Thr Pro Gly Tyr Ile Phe Gly Lys Arg Ile Ile Leu Phe Leu Phe Leu
    110                 115                 120
atg tcc gtt gct ggc ata ttc aac tat tac ctc atc ttc ttt ttc gga      494
Met Ser Val Ala Gly Ile Phe Asn Tyr Tyr Leu Ile Phe Phe Phe Gly
125                 130                 135                 140
agt gac ttt gaa aac tac ata aag acg atc tcc acc acc atc tcc cct      542
Ser Asp Phe Glu Asn Tyr Ile Lys Thr Ile Ser Thr Thr Ile Ser Pro
                145                 150                 155
cta ctt ctc att ccc taa ctctctgctg aatatggggt tggtgttctc             590
Leu Leu Leu Ile Pro  *
            160
atctaatcaa tacctacaag tcatcataat tcagctcttg agagcattct gctcttcttt    650
agatggctgt aaatctattg gccatctggg cttcacagct tgagttaacc ttgcttttcc    710
gggaacaaaa tgatgtcatg tcagctccgc cccttgaaca tgaccgtggc cccaaatttg    770
ctattcccat gcattttgtt tgtttcttca cttatcctgt tctctgaaga tgttttgtga    830
ccaggtttgt gttttcttaa aataaaatgc agagacatgt tttaa                    875
```

<210> 3
<211> 161
<212> PRT
<213> homo sapiens

<220>
<221> VARIANT
<222> 127
<223> 10-36-164 : polymorphic amino acid Val or Ile

<400> 3
Met Asp Gln Glu Thr Val Gly Asn Val Val Leu Leu Ala Ile Val Thr
1 5 10 15
Leu Ile Ser Val Val Gln Asn Gly Phe Phe Ala His Lys Val Glu His
20 25 30
Glu Ser Arg Thr Gln Asn Gly Arg Ser Phe Gln Arg Thr Gly Thr Leu
35 40 45
Ala Phe Glu Arg Val Tyr Thr Ala Asn Gln Asn Cys Val Asp Ala Tyr
50 55 60
Pro Thr Phe Leu Ala Val Leu Trp Ser Ala Gly Leu Leu Cys Ser Gln
65 70 75 80
Val Pro Ala Ala Phe Ala Gly Leu Met Tyr Leu Phe Val Arg Gln Lys
85 90 95
Tyr Phe Val Gly Tyr Leu Gly Glu Arg Thr Gln Ser Thr Pro Gly Tyr
100 105 110
Ile Phe Gly Lys Arg Ile Ile Leu Phe Leu Phe Leu Met Ser Val Ala
115 120 125
Gly Ile Phe Asn Tyr Tyr Leu Ile Phe Phe Phe Gly Ser Asp Phe Glu
130 135 140
Asn Tyr Ile Lys Thr Ile Ser Thr Thr Ile Ser Pro Leu Leu Leu Ile
145 150 155 160
Pro

<210> 4
<211> 46
<212> DNA
<213> homo sapiens
<400> 4
ctacactgcc aagtgagtcc taaacctgat gttgctaata agtggg 46

<210> 5
<211> 46
<212> DNA
<213> homo sapiens
<400> 5
ctacactgcc aagtgagtcc taaccctgat gttgctaata agtggg 46

<210> 6
<211> 19
<212> DNA
<213> homo sapiens
<400> 6
ggacatttag ggttgcttg 19

<210> 7
<211> 19
<212> DNA
<213> homo sapiens
<400> 7
tgttttgaca ctagcactc 19

<210> 8
<211> 19

<212> DNA
<213> homo sapiens
<400> 8
actgccaagt gagtcctaa 19

<210> 9
<211> 48
<212> DNA
<213> homo sapiens
<400> 9
ctgactagga cctctgattc ttctctccct gagctttgaa ggctctga 48

<210> 10
<211> 48
<212> DNA
<213> homo sapiens
<400> 10
ctgactagga cctctgattc ttctttccct gagctttgaa ggctctga 48

<210> 11
<211> 18
<212> DNA
<213> homo sapiens
<400> 11
gaaggttctc aggtttcc 18

<210> 12
<211> 18
<212> DNA
<213> homo sapiens
<400> 12
agctgtattt tcagagcc 18

<210> 13
<211> 19
<212> DNA
<213> homo sapiens
<400> 13
taggacctct gattcttct 19

<210> 14
<211> 18
<212> DNA
<213> Homo Sapiens

<220>
<221> misc_binding
<222> 1..18
<223> sequencing oligonucleotide PrimerPU

<400> 14
tgtaaaacga cggccagt 18

<210> 15
<211> 18
<212> DNA
<213> Homo Sapiens

<220>
<221> misc_binding
<222> 1..18
<223> sequencing oligonucleotide PrimerRP

<400> 15
caggaaacag ctatgacc 18

Evidence Appendix C

Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4846–4850, July 1987
Biochemistry

Express Mail
Mailing Label No. EV832482014GS

Complete cDNA and derived amino acid sequence of human factor V

(cDNA cloning/sequence homology/blood coagulation/gene evolution)

RICHARD J. JENNY*, DEBRA D. PITTMAN‡, JOHN J. TOOLE‡, RONALD W. KRIZ‡, ROBERT A. ALDAPE‡, RODNEY M. HEWICK‡, RANDAL J. KAUFMAN‡, AND KENNETH G. MANN*†

*Department of Biochemistry, College of Medicine, University of Vermont, Burlington, VT 05405; and ‡Genetics Institute, Cambridge, MA 02140

Communicated by Russell F. Doolittle, March 27, 1987

ABSTRACT **cDNA clones encoding human factor V have been isolated from an oligo(dT)-primed human fetal liver cDNA library prepared with vector Charon 21A. The cDNA sequence of factor V from three overlapping clones includes a 6672-base-pair (bp) coding region, a 90-bp 5′ untranslated region, and a 163-bp 3′ untranslated region within which is a poly(A) tail. The deduced amino acid sequence consists of 2224 amino acids inclusive of a 28-amino acid leader peptide. Direct comparison with human factor VIII reveals considerable homology between proteins in amino acid sequence and domain structure: a triplicated A domain and duplicated C domain show ≈40% identity with the corresponding domains in factor VIII. As in factor VIII, the A domains of factor V share ≈40% amino acid-sequence homology with the three highly conserved domains in ceruloplasmin. The B domain of factor V contains 35 tandem and ≈9 additional semiconserved repeats of nine amino acids of the form Asp-Leu-Ser-Gln-Thr-Thr/Asn-Leu-Ser-Pro and 2 additional semiconserved repeats of 17 amino acids. Factor V contains 37 potential N-linked glycosylation sites, 25 of which are in the B domain, and a total of 19 cysteine residues.**

Factor V is a large and asymmetric glycoprotein that circulates in plasma and is an essential component of the blood coagulation cascade (1, 2). During coagulation, the procofactor factor V is converted to the active cofactor, factor Va, via limited proteolysis by α-thrombin (3–5). Factor Va is a cofactor for the serine protease factor Xa, and together, factors Va and Xa assemble on a cellular or phospholipid surface with divalent metal ions to form the prothrombinase complex (1, 6–11). This complex enhances factor Xa activity ≈350,000 fold. The prothrombinase complex is analogous to another complex that proteolytically cleaves zymogen factor X to active enzyme factor Xa—this other "ten-ase" complex is composed of a serine protease (factor IXa), a cofactor (factor VIIIa), phospholipid, and calcium (13, 14).

In addition to the similarities between serine proteases (factors Xa and IXa) and in overall enzyme complex architecture, the cofactors (factor Va and factor VIIIa) are very similar proteins structurally and functionally (13–16). Heavy and light chains of bovine factor Va and porcine factor VIIIa possess amino acid-sequence homology at the amino-terminal portion of the chains—regions of homology that are also homologous to regions in the triplicated domain structure of ceruloplasmin, the primary transport protein for copper in plasma. Available data therefore suggest that factor V, factor VIII, and ceruloplasmin are members of a family of structurally related proteins (15).

The molecular cloning and sequencing of human factor VIII and human ceruloplasmin gives evidence for a common domain structure and has enabled detailed comparison of their structures (16, 17–20). Recently Kane and Davie (21) published a partial cDNA sequence for human factor V that coded for ≈40% of the molecule. This cDNA coded for the light chain and a small portion of the heavily glycosylated connecting region. We present the complete cDNA and deduced amino acid sequence of human factor V and compare this sequence with the primary structures of factor VIII and ceruloplasmin.

MATERIALS AND METHODS

Materials and Reagents. All DNA-modifying enzymes and cloning vectors were obtained from either New England Biolabs, Bethesda Research Laboratories, or Promega Biotec; all other reagents and supplies were of high quality and are commercially available.

Screening of Human Fetal Liver cDNA Libraries. Using amino acid sequence from the amino-terminal portion of human factor Va light chain (15, 22), we synthesized a 39-mer oligonucleotide (AACTAYTAYATTGCTGCTGAGGAGATCACCTGGGACTAT, where: Y = T or C) on an Applied Biosystems DNA synthesizer and subsequently end-labeled it with T4 polynucleotide kinase and [γ^{32}-P]ATP (ref. 23, pp. 122–123). This probe was then used to screen an oligo(dT)-primed human fetal liver Charon 21A library (16); clone V1 obtained on the initial screening was used to generate two new probes. Restriction fragments obtained were labeled with [α-^{32}P]ATP by nick translation (ref. 23, pp. 109–112) or random priming (24) and were used to screen for additional clones.

Isolation, Subcloning, and Sequencing of Insert cDNA. Phage DNA was prepared from positive clones and digested with appropriate restriction enzymes to isolate cDNA inserts. Inserts were subcloned into the plasmid vector SP65 (Promega Biotec) or M13 phage cloning vector mp18 or mp19 (New England Biolabs) as described by Maniatis *et al.* (ref. 23, pp. 390–401). Nucleotide sequence analysis was done as described by Poncz *et al.* (25), Messing and Vieira (26), and Sanger *et al.* (27). cDNA and protein homologies were studied by computer-assisted searches of Genbank and the National Biomedical Research Foundation protein sequence data base.

RNA and Southern Blot Analysis. RNA blot analysis was used to estimate the size of the mRNA and the relative abundance of the message in various cell types and to help confirm full-length clones. mRNA was isolated from cultured cells using the guanidine thiocyanate method (28), and both mRNA and poly(A)$^+$ mRNA were isolated from human liver as described (29). mRNA was fractionated on a 0.8% agarose-formaldehyde gel and was subsequently transferred to nitrocellulose as described (30). The 1614-bp *Eco*RI restriction

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.

enzyme fragment was isolated from clone V1 and was used to probe the RNA blots. Similar analysis with other restriction fragments verified factor V clones. Common restriction enzyme fragments in other clones were identified by Southern blot analysis, which was done as described (ref. 23, pp. 383–386).

RESULTS AND DISCUSSION

Approximately 10^5 recombinant λ phage plaques from a human fetal liver cDNA library were screened with the 39-mer oligonucleotide probe derived from amino acid sequencing of the human factor Va light chain, and 14 positive clones were identified. A phage designated V1 was chosen for further analysis. cDNA sequence encoding the amino acids from which the 39-mer was designed confirmed that the clone contained factor V DNA. The cDNA insert in phage V1 was too short to encode the entire factor V molecule. To obtain cDNA encoding the amino terminus of factor V, restriction enzyme fragments from the 5′ end of clone V1 were isolated, labeled with ^{32}P, and used to screen additional recombinant λ phage. Approximately 10^6 recombinant phage were screened, and 92 positive clones were identified; two phage designated V401 and V402 were selected for further analysis. Phage V401 contained an *Eco*RI fragment common to both clones as shown by cross-hybridization analysis, but which was longer than the equivalent fragment in phage V402 (as determined by restriction enzyme digestion and subsequent fractionation in agarose gels). Further analysis of this fragment revealed an open reading frame with cDNA sequence encoding the amino-terminal portion of the intact factor V molecule, thus confirming a clone that encoded amino acid sequence beyond the amino terminus of the mature plasma protein.

A schematic representation of the three overlapping clones that were used to complete the cDNA sequence of factor V is shown (Fig. 1); Fig. 2 illustrates cDNA and the deduced amino acid sequence. The cDNA sequence includes a 6672-bp coding region, a 90-bp 5′ untranslated region, and a 163-bp 3′ untranslated region including a poly(A) tail. The coding region begins at nucleotide 91 with the initiator codon ATG. The 3′ untranslated region contains the putative polyadenylylation signal sequence AATAAA (31), located 12 nucleotides 5′ of the poly(A) tail. The deduced amino acid sequence of factor V consists of 2224 amino acids that include a 28-amino acid leader peptide. The leader peptide contains a cluster of hydrophobic amino acids and an alanine residue at position -1, which is consistent with the structure of most leader peptides and with known specificities of cleavage by leader peptidases (32–34).

Size and abundance of the factor V mRNA from various tissue sources were analyzed by RNA blot hybridization. The results (Fig. 3) indicate that a factor V message of ≈7000 bp is quite abundant in both liver and HepG2 cells. Prolonged exposure of the RNA blots did not indicate any factor V mRNA contained within the RNA isolated from human umbilical-vein endothelial cells, peripheral blood leukocytes, or U297 cells.

The complete sequence illustrated in Fig. 2 verifies and extends the partial cDNA sequence encoding the 3′ half of factor V recently published by Kane and Davie (21). These data indicate that factor V has a domain structure similar to that of factor VIII and that these molecules share ≈40% amino acid sequence identity in the A and C domains. The domains within factor V are schematically illustrated in Fig. 1 and correspond to the following regions: A1, residues 1–317; A2, residues 318–663; A3, residues 1546–1883; B, residues 664–1545; C1, residues 1884–2036; and C2, residues 2037–2196.

A comparison of factor V, factor VIII, and ceruloplasmin by the Dayhoff protein alignment method (35) is represented in Fig. 4. These data indicate highly conserved domains within and between these proteins. In addition, it has also been pointed out that the C domains of factor V and factor VIII share a 20% amino acid-sequence homology with the discoidins, which are phospholipid-binding lectins from *Dictyostelium Discoideum* (21, 36). The B domains of factors V and VIII share a 14% amino acid sequence identity, in

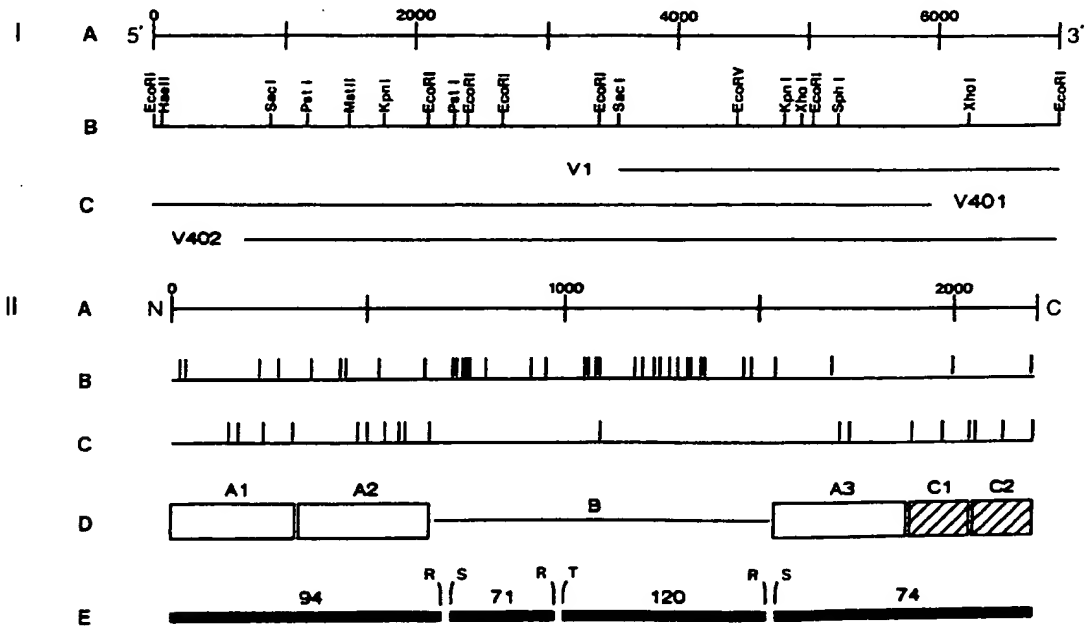

FIG. 1. (*I*) Line diagram of human factor V cDNA: (*A*) cDNA number line, (*B*) partial restriction enzyme map, and (*C*) three overlapping cDNA clones used to obtain the complete cDNA sequence. (*II*) Line diagram of human factor V protein: (*A*) amino acid number line, N, amino terminus, C, carboxyl terminus, (*B*) potential N-linked glycosylation sites, (*C*) cysteine residues, (*D*) domain structure based upon internal homologies and homologies with factor VIII and ceruloplasmin, and (*E*) thrombin cleavage products labeled according to their putative M_r values with carbohydrate included. R, Arg; S, Ser; and T, Thr.

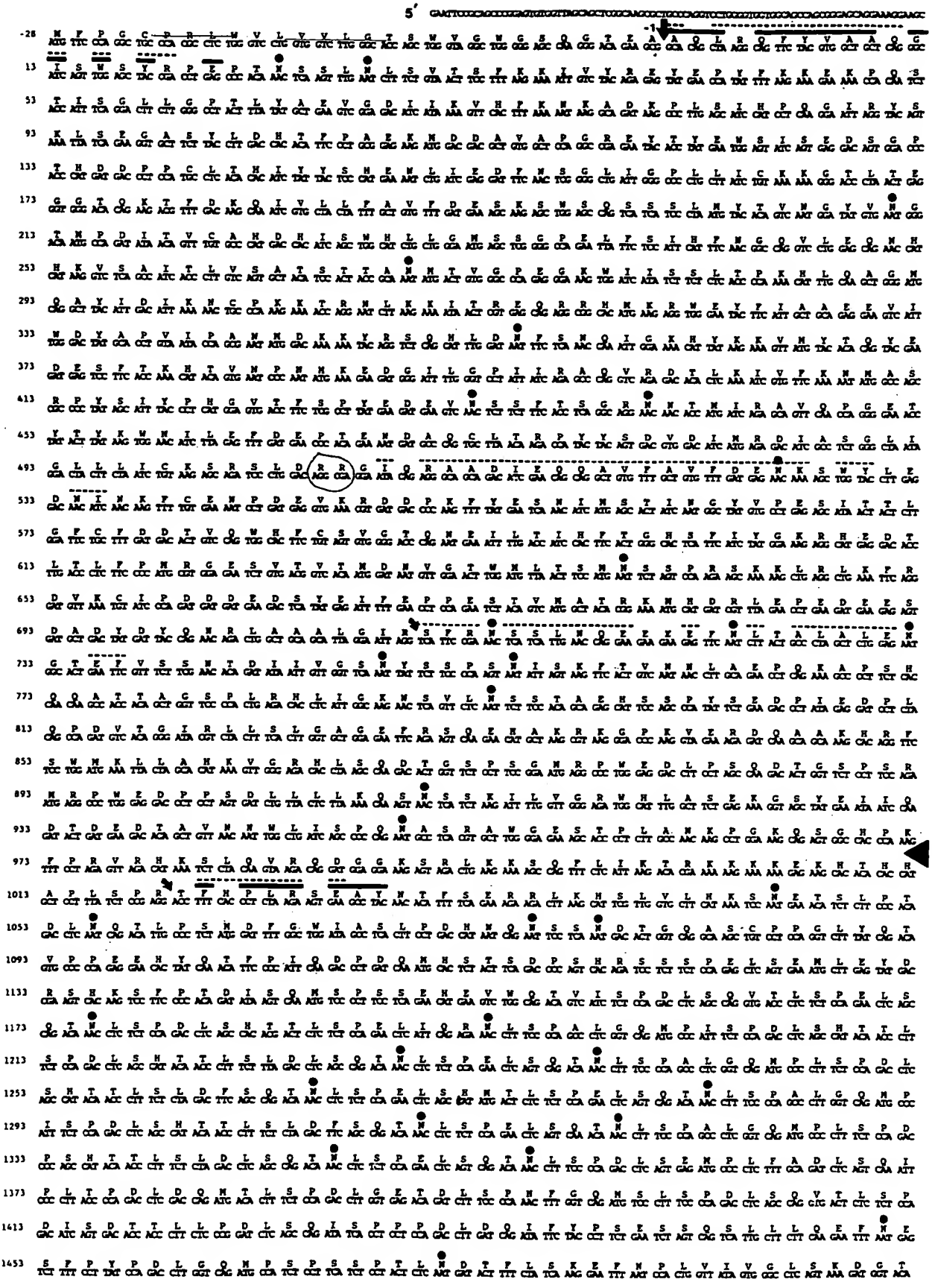

FIG. 2. (*Figure continues on the opposite page.*)

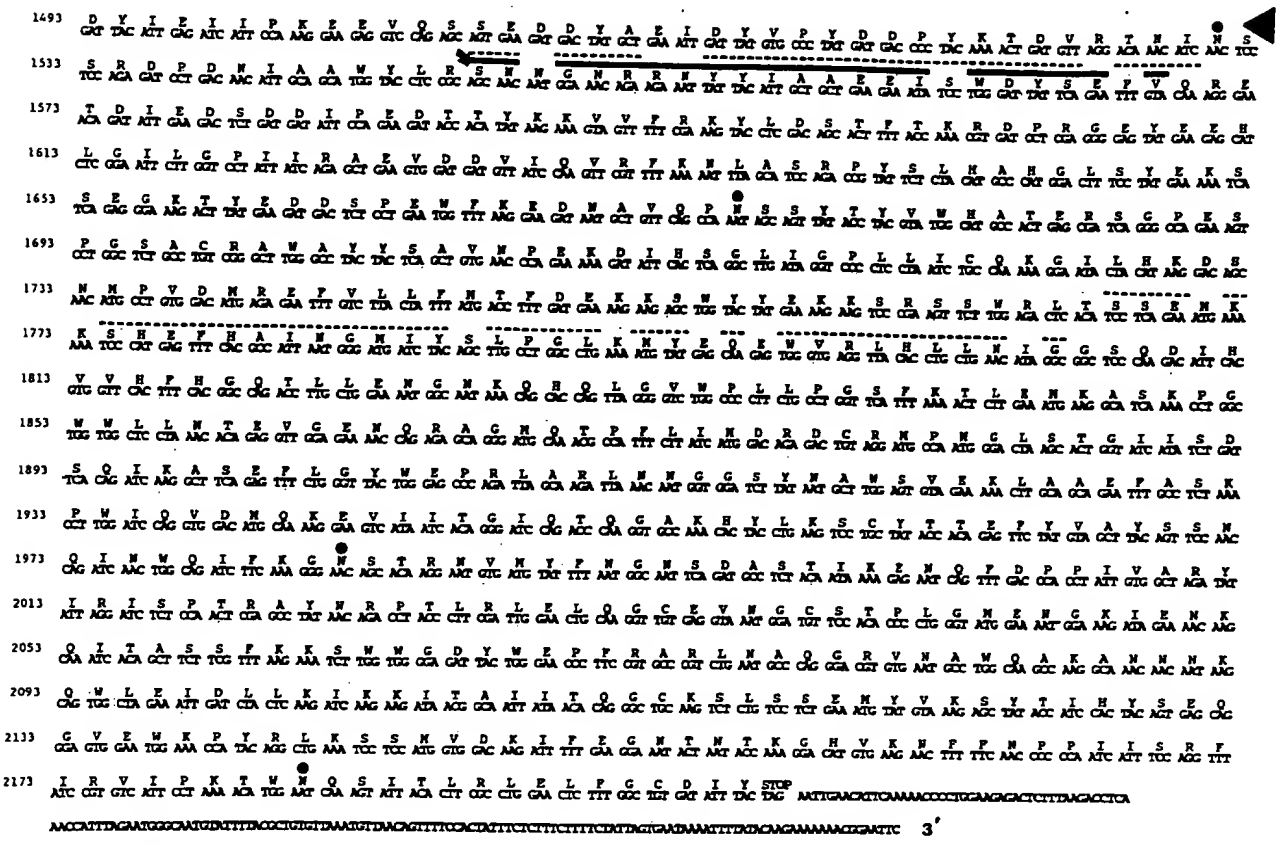

FIG. 2. Complete cDNA sequence and deduced amino acid sequence of human factor V. Sequence is numbered by amino acids with position 1 corresponding to the amino-terminal residue of the plasma protein. Heavy vertical arrow, leader peptide cleavage site; filled circles, potential N-linked glycosylation sites; curved arrows, thrombin cleavage sites. Amino acid sequence determined by amino acid sequencing is indicated by the solid overbar for sequence obtained from human factor V and by a dashed overbar for sequence obtained from bovine factor V. The region containing the 44 semiconserved repeats of nine amino acids is bordered by the solid triangles at the right.

contrast to an ≈40% identity shared by the other domains (Fig. 4). A search of the National Biomedical Research Foundation protein sequence data bank revealed significant homologies with only factor VIII, ceruloplasmin, and discoidin I.

The B domain of factor V is structurally unique relative to the rest of the molecule. Factor V contains 37 potential N-glycosylation sites of the form Asn-Xaa-Ser/Thr of which 25 are located in the B domain (Fig. 1)—consistent with the reports of heavily glycosylated activation products (5, 37–39). Because of thrombin activation of factor V to factor Va, two peptides of estimated M_r 71,000 and 120,000 are released as activation peptides (3–5). From amino acid composition we have determined the M_r (excluding carbohydrate) of these peptides as ≈34,000 and 58,000, respectively. With the M_r values estimated for these peptides, the data indicate a carbohydrate composition of ≈50% by weight. Factor V contains 19 cysteine residues, only one of which is located in the B domain in the 120,000 fragment (Fig. 1). The products from thrombin activation of factor V are not covalently associated, thus indicating a free sulfhydryl group in the 120,000 fragment. The B domain of factor V contains a region of 35 tandem and ≈9 additional semiconserved repeats of 9 amino acids of the form Asp-Leu-Ser-Gln-Thr-Thr/Asn-Leu-Ser-Pro (D-L-S-Q-T-T/N-L-S-P) as shown in Fig. 2. This is larger than the 20 repeats seen by Kane and Davie (21) and may be due to a 297-bp in-frame deletion within this repetitive region. In addition to these repeated sequences, the B domain also contains two highly conserved repeats of 17 amino acids upstream from the tandem repeat region.

A highly reactive free sulfhydryl group and the clustered repeats that contain Glu/Asp residues may contribute to the capability of the 120,000 fragment as a substrate for the

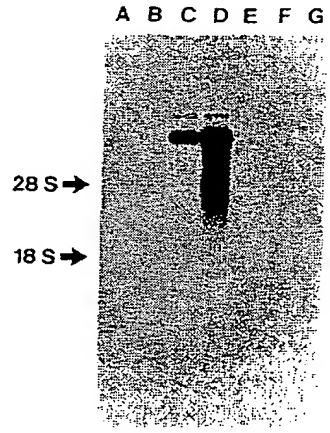

FIG. 3. RNA blot analysis of factor V mRNA. The RNA blot shown was hybridized with the 1614-bp *Eco*RI fragment of clone V1 as described. The 28S and 18S ribosomal RNA band positions are included for reference. (Lane A) human liver total RNA (5 μg); (lane B) human umbilical-vein endothelial cell total RNA (5 μg); (lane C) human HepG2 cell total RNA (5 μg); (lane D) human liver poly(A)$^+$ selected RNA (4 μg); (lane E) human liver poly(A)$^+$ selected RNA (2 μg); (lane F) human peripheral blood leukocyte total RNA (2 μg); and (lane G) human U297 (transformed monocyte) cell total RNA (2 μg).

MATCHES/LENGTH (%)

	V-A2	V-A3	VIII-A1	VIII-A2	VIII-A3	CER-A1	CER-A2	CER-A3	V-C2	VIII-C1	VIII-C2	VIII-B
V-A1	31	32	36	30	27	31	32	21				
V-A2		29	29	44	31	35	34	35				
V-A3			33	31	39	32	34	39				
VIII-A1				30	32	34	32	33				
VIII-A2					33	38	35	36				
VIII-A3						28	34	36				
CER-A1							37	39				
CER-A2								40				
V-C1									37	44	34	
V-C2										41	42	
VIII-C1											38	
V-B												14

FIG. 4. Comparison of the homologous domains in factor V, factor VIII, and ceruloplasmin. Values in the table represent the total identical amino acid matches divided by the overlapping lengths (including gaps) expressed as percentages.

transglutaminase activity of factor XIIIa seen by Francis *et al.* (40). In addition, modification of these reactive groups may produce effects that will elucidate the function of B domain. The B domain of factor VIII has been mapped to a single exon of ≈3000 bp (19). From current theories of gene evolution and RNA processing, one would predict multiple exons for the repeat regions of the factor V B domain.

Activation of factor V to factor Va involves three specific enzymatic cleavages catalyzed by thrombin (3–5). Amino acid sequence from α-thrombin cleavage products shows thrombin cleavage sites at the following positions: Arg-709/Ser-710, Arg-1018/Thr-1019, and Arg-1545/Ser-1546. A schematic of the thrombin activation products is shown in Fig. 1, and the cleavage sites are further illustrated in Fig. 2.

The primary structure of factor V will contribute to better understanding the functions of regions in factor V and factor VIII in membrane binding and interaction with serine proteases factor Xa and factor IXa. By comparative analysis of structural domains and site-directed mutagenesis, functionality of regions in factor V and factor VIII can be detailed. In addition, much is to be learned about gene evolution within the family of ceruloplasmin-related proteins.

The authors thank John Knopf and Lisa Sultzman for helpful discussions and Darlene Vanstone and John Brown for oligonucleotide synthesis. This work was supported in part by National Institutes of Health Grant HL-34575.

1. Nesheim, M. E., Katzman, J. A., Tracy, P. B. & Mann, K. G. (1981) *Methods Enzymol.* **80**, 249–274.
2. Jackson, C. M. & Nemerson, Y. (1980) *Annu. Rev. Biochem.* **49**, 765–811.
3. Esmon, C. T. (1979) *J. Biol. Chem.* **254**, 964–973.
4. Nesheim, M. E. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 1326–1334.
5. Suzuki, K., Dahlback, B. & Stenflo, J. (1982) *J. Biol. Chem.* **257**, 6556–6564.
6. Miletich, J. P., Jackson, C. M. & Majerus, P. W. (1978) *J. Biol. Chem.* **253**, 6908–6916.
7. Dahlback, B. & Stenflo, J. (1978) *Biochemistry* **17**, 4938–4945.
8. Tracy, P. B., Nesheim, M. E. & Mann, K. G. (1981) *J. Biol. Chem.* **256**, 743–751.
9. Tracy, P. B., Eide, L. L. & Mann, K. G. (1985) *J. Biol. Chem.* **260**, 2119–2124.
10. Nesheim, M. E., Taswell, J. B. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 10952–10962.
11. Nesheim, M. E., Kettner, C., Shaw, E. & Mann, K. G. (1981) *J. Biol. Chem.* **256**, 6537–6540.
12. Krishnaswamy, S., Church, W. R., Nesheim, M. E. & Mann, K. G. (1987) *J. Biol. Chem.* **262**, 3291–3299.
13. Mann, K. & Fass, D. (1983) in *Hematology*, ed. Fairbanks, V. F. (Wiley, New York), Vol. 2, pp. 347–374.
14. Mann, K. G. (1984) in *Progress in Hemostasis and Thrombosis*, ed. Spaet, T. H. (Grune & Stratton, Orlando, FL), Vol. 7, pp. 1–23.
15. Church, W. R., Jernigan, R. L., Toole, J., Hewick, R. M., Knopf, J., Knutson, G. J., Nesheim, M. E., Mann, K. G. & Fass, D. N. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6934–6937.
16. Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, W. B., Coe, M. L., Knutson, G. J., Fass, D. N. & Hewick, R. M. (1981) *Nature (London)* **312**, 342–347.
17. Takahashi, N., Ortel, T. L. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 390–394.
18. Koschinsky, M. L., Funk, W. D., van Oost, B. A. & MacGillivray, R. T. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5086–5090.
19. Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J. & Lawn, R. M. (1984) *Nature (London)* **312**, 326–330.
20. Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkins, R. N., Tuddenham, E. G. D., Lawn, R. M. & Capon, D. J. (1984) *Nature (London)* **312**, 337–342.
21. Kane, W. H. & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6800–6804.
22. Katzman, J. A., Nesheim, M. E., Hibbard, L. S. & Mann, K. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 162–166.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
24. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13, and addendum (1984) **137**, 266–267.
25. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4298–4302.
26. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
28. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
29. Hastie, N. D., Held, W. A. & Toole, J. J. (1979) *Cell* **17**, 449–457.
30. Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M. & Darnell, J. E. (1981) *Cell* **23**, 731–739.
31. Proudfoot, N. H. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
32. von Heijne, G. (1982) *J. Mol. Biol.* **159**, 537–541.
33. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164.
34. von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
35. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
36. Poole, S., Firtel, R. A., Lamar, E. & Rowekamp, W. (1981) *J. Mol. Biol.* **153**, 273–289.
37. Kane, W. H. & Majerus, P. W. (1981) *J. Biol. Chem.* **256**, 1002–1007.
38. Esmon, C. T. (1980) in *The Regulation of Coagulation*, ed. Mann, K. G. & Taylor, F. B. (Elsevier, New York), pp. 137–143.
39. Lindhout, T., Grovers-Riemslag, J. W. P., van de Waart, P., Hemker, H. C. & Rosing, J. (1982) *Biochemistry* **21**, 5494–5502.
40. Francis, R. T., McDonagh, J. & Mann, K. G. (1986) *J. Biol. Chem.* **261**, 9787–9792.

Evidence Appendix D

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6800–6804, September 1986
Biochemistry

Express Mail
Mailing Label No. EV832482014US

Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin

(cDNA cloning/DNA sequence analysis/blood coagulation)

WILLIAM H. KANE AND EARL W. DAVIE

Department of Biochemistry, University of Washington, Seattle, WA 98195

Contributed by Earl W. Davie, June 12, 1986

ABSTRACT **Coagulation factor V is a high molecular weight plasma glycoprotein that participates as a cofactor in the conversion of prothrombin to thrombin by factor X_a. A phage λgt11 Hep G2 cell cDNA expression library was screened by using an affinity-purified antibody to human factor V, and 11 positive clones were isolated and plaque-purified. The clone containing the largest cDNA insert contained 2970 nucleotides and coded for 938 amino acids, a stop codon, and 155 nucleotides of 3′ noncoding sequence including a poly(A) tail. The coding region includes 651 amino acids from the carboxyl terminus that constitute the light chain of human factor V_a and 287 amino acids that are part of the connecting region of the protein. The predicted amino acid sequence agreed completely with 147 amino acid residues that were identified by Edman degradation of cyanogen bromide peptides isolated from the light chain. During the activation of factor V, several peptide bonds are cleaved by thrombin, giving rise to a heavy chain, a connecting fragment(s), and a light chain. The light chain is generated by the cleavage of an Arg-Ser peptide bond. The amino acid sequence of the light chain is homologous (40%) with the carboxyl-terminal fragment (M_r, 73,000) of human factor VIII. Both fragments have a similar domain structure that includes a single ceruloplasmin-related domain followed by two C domains. The carboxyl terminus of the connecting region, however, shows no significant amino acid sequence homology with factor VIII. It is very acidic and contains a number of potential N-linked glycosylation sites. It also contains about 20 tandem repeats of nine amino acids.**

Human coagulation factor V is a high molecular weight plasma glycoprotein that is required for rapid thrombin formation and normal hemostasis (1). It circulates in blood as a large single polypeptide chain (M_r, 330,000) with little or no coagulant activity (2–6). During the blood coagulation process, factor V is converted to factor V_a by thrombin by limited proteolysis (2–6). This makes available binding sites for factor X_a (7) and prothrombin (8). Factor V_a is composed of a heavy chain (M_r, 110,000) and a light chain (M_r, 76,000), and these two chains are held together by calcium ions (3, 6). The remainder of the original factor V molecule is released as a large connecting fragment(s) that is rich in carbohydrate (3, 6, 9). Factor V_a binds to cell surfaces (9–12) and negatively charged phospholipid surfaces (13–16) through the light chain, and this increases the rate of prothrombin activation ≈10,000-fold by factor X_a (17). Factor V_a is readily inactivated by activated protein C (18), and this results in the cleavage of the heavy chain into two smaller fragments (19). In addition to its role in prothrombin activation, human factor V_a stimulates the activation of protein C by thrombin on phospholipid and cell surfaces (12, 20, 21). Only the light chain is necessary for this reaction (22).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Factor V has a number of physical and biological properties that are similar to factor VIII. For instance, factor VIII is also converted to an activated form (factor $VIII_a$) by thrombin, and this molecule enhances the rate of activation of factor X by factor IX_a. This cofactor effect of factor $VIII_a$ is analogous to that of factor V_a on the activation of prothrombin. Fass *et al.* (23) have also reported amino acid sequence homology between the N-terminal portions of the heavy and light chains of bovine factor V_a and the corresponding peptides from porcine factor $VIII_a$. In addition, these sequences showed amino acid sequence homology with ceruloplasmin, a plasma copper-binding protein (24). Subsequently, both bovine and human factor V have been shown to contain one copper ion per molecule (25). The complete amino acid sequence of human factor VIII has been determined by cDNA cloning (26, 27). It contains a triplicated A domain (≈350 amino acids) with ≈30% internal amino acid homology. These A domains are also ≈30% homologous with the triplicated domains of ceruloplasmin (28). In contrast to ceruloplasmin, factor VIII also contains a connecting peptide (≈900 amino acids) located between the second and third A domains and contains two C domains (≈150 amino acids) located near the carboxyl-terminal end of the molecule.

The biosynthesis of factor V has been demonstrated in human Hep G2 cells (29), bovine aortic endothelial cells (30), and guinea pig megakaryocytes (31). In this manuscript, we report the isolation of a cDNA clone from a Hep G2 cDNA library that codes for the carboxyl-terminal 938 amino acids of human factor V. This includes the entire light chain of factor V_a (651 amino acids) and a portion of the connecting region (287 amino acids).

MATERIALS AND METHODS

Screening of the λgt11 cDNA Library. Human factor V was prepared by the method of Kane and Majerus (4), and rabbit antibodies were prepared and affinity-purified by the method of Canfield and Kisiel (32). The Hep G2 λgt11 cDNA expression library was kindly provided by Frederick S. Hagen (33). The affinity-purified antibody to human factor V was labeled with ^{125}I and then used to screen filter blots as described by Young and Davis (34). Positive clones were also hybridized with a ^{32}P-labeled oligonucleotide probe with a sequence of ACCCAYTCYTGYTCRTACAT, where Y is T or C and R is G or A (35). This oligonucleotide was synthesized by using an Applied Biosystems (Foster City, CA) oligonucleotide synthesizer.

DNA Sequence Analysis. Phage DNA was prepared from positive clones (36, 37). The cDNA inserts were isolated after *Eco*RI digestion. The inserts were then subcloned into pUC9 (38), followed by subcloning of appropriate restriction fragments into M13 phage cloning vectors mp18 and mp19. DNA sequencing was performed by the BAL-31 exonuclease method (39). Dideoxy chain termination sequencing reactions were carried out with ^{35}S-substituted deoxyadenosine 5′-[α-thio]triphosphate (Amersham) and universal M13 prim-

ers as described (39). DNA sequences were stored and analyzed on an Apple Macintosh computer using the DNA INSPECTOR program (Textco, West Lebanon, VT) (40).

Protein Sequence Analysis. The light chain of factor V_a was isolated from 10 mg of human factor V after incubation with the factor V activator from Russell's viper venom (6). The light chain was then reduced, carboxymethylated (41), and digested with cyanogen bromide (42). The cyanogen bromide peptides were purified by gel filtration on Sephadex G-50 followed by HPLC using an Altex Ultrapore C3 reverse-phase separation column as described by McMullen and Fujikawa (41). Seven peptides were isolated in sufficient yield to allow automated amino acid sequence analysis with a Beckman 890C sequencer. Amino acid sequences were analyzed by using the ALIGN computer program of Dayhoff *et al.* (43).

RESULTS AND DISCUSSION

Affinity-purified antibody to human factor V was radiolabeled with ^{125}I and used for the screening of $\approx 2 \times 10^6$ phage from a λgt11 Hep G2 cell cDNA expression library. Eleven positive clones were identified and plaque-purified. The clone containing the largest cDNA insert (λHV2970) also hybridized with ACCCAYTCYTGYTCRTACAT, an oligonucleotide complementary to the DNA that codes for the amino acid sequence of Met-Tyr-Glu-Gln-Glu-Trp-Val. This amino acid sequence is present in the light chain of human factor V_a. The cDNA insert was then analyzed by restriction mapping and sequenced two or more times (Fig. 1). Also, ≈95% of the sequence was determined on both strands. The cDNA insert contained 2970 nucleotides and coded for 938 amino acids, a stop codon (TAG), and 155 nucleotides of 3′ noncoding sequence, including a polyadenylylation signal (AATAAA) and a poly(A) tail (Fig. 2). The predicted amino acid sequence was in complete agreement with 147 amino acids that were identified by Edman degradation of seven cyanogen bromide peptides. These peptides were isolated and purified from the light chain of factor V_a and are overlined in Fig. 2. These data indicate that the cDNA insert in λHV2970 coded for the carboxyl terminus of the protein, including 651 amino acids that constitute the entire light chain of factor V_a (see below) and 287 amino acids that form a portion of the connecting region.

The molecular weight of the light chain of human factor V_a was calculated to be 71,500 without carbohydrate. The addition of the three potential N-linked oligosaccharide chains with a molecular weight of ≈2000 each increases the molecular weight of the glycoprotein to ≈77,500. This is similar to the values of 74,000–78,000, as determined by $NaDodSO_4$/polyacrylamide gel electrophoresis (4–6). The amino acid sequence of the light chain is 40% homologous with the carboxyl-terminal fragment (M_r, 73,000) of human factor VIII (26, 27). Both fragments have a similar domain structure, including a single ceruloplasmin-related A domain followed by two C domains. The A domain in the light chain of factor V_a is ≈40% homologous with the third A domain of both human ceruloplasmin and human factor VIII (Fig. 3*A*). The A domain of the light chain of factor V_a, however, shows

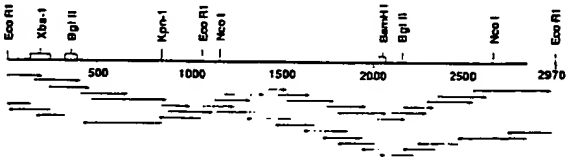

FIG. 1. Partial restriction map and sequencing strategy for the cDNA insert in λHV2970 that codes for human factor V. The extent of sequencing is shown by the length of each arrow, and the direction of the arrow indicates the strand that was sequenced. The coding region is indicated by the solid black line.

less sequence homology (30–36%) with the first and second A domains of factor VIII and ceruloplasmin. Also, the A domain of the light chain of factor V_a contains only a single pair of cysteine residues, whereas the corresponding domains in human factor VIII and ceruloplasmin contain five and three cysteine residues, respectively.

The location and function of the single copper ion associated with factor V remain unknown. Ceruloplasmin contains six copper ions in three types of binding sites (44). Mann *et al.* (25) could not detect type I (blue, 610-nm absorbance) or type III (310-nm absorbance) binding sites in factor V and concluded that the copper may be bound to a type II binding site. By analogy with the type I copper-binding protein plastocyanin, the ligands for the type I copper-binding site at the carboxyl terminus of ceruloplasmin have been proposed to be His-975, Cys-1021, His-1026, and Met-1031 (45). Only two of these four residues are conserved in the light chain of factor V_a (see Fig. 3*A*). This is in contrast to factor VIII, in which all four residues are conserved in the first and third A domains (26, 27). The carboxyl terminus of ceruloplasmin also contains eight closely clustered histidine residues, which have been suggested to play a role in copper binding (45). Of these eight histidine residues, only three are conserved in factor V (Fig. 2), while five are conserved in factor VIII (26).

The two C domains of human factor V show 35–50% homology with each other and with the C domains of human factor VIII (Fig. 3*B*). As was noted previously in factor VIII (29), these domains share 20% homology with the slime mold protein discoidin I (46). Discoidin I is a tetrameric galactose-binding lectin, which is essential for cell adhesion in *Dictyostelium discoideum*. It contains the amino acid sequence Arg-Gly-Asp that interacts with a specific cell surface receptor (47). The Arg-Gly-Asp sequence is not present in the two C domains of factor V or factor VIII. The light chain of factor V_a binds to negatively charged phospholipid via electrostatic interactions (14, 16). Pusey and Nelsestuen (16) found that binding was inhibited by 90% when 12% of the lysine residues were modified by citraconic anhydride. It is notable that the C domains in the light chain of factor V_a are basic with a calculated charge at pH 7.0 of +14 (excluding carbohydrate). There are several clusters of basic residues, including a cluster of four or five arginine and lysine residues at the end of each C domain. In contrast, the A domain of the light chain of factor V_a has a calculated charge of −9. A similar charge distribution is seen in the corresponding regions of human factor VIII. It is possible that the binding of the light chain of factor V_a to cell and phospholipid surfaces is mediated by the C domains. However, the precise functions of the various domains in factor V must await future experiments.

During the activation of factor V, the light chain is generated by the cleavage of an Arg-Ser bond located between the connecting region and the light chain. Factor V is rapidly activated by thrombin or the factor V activator in Russell's viper venom. The amino acid sequences determined by Edman degradation of the light chain of factor V_a generated by thrombin (D. B. Wilson, personal communication) or the factor V activator in Russell's viper venom (Fig. 2) are identical, indicating that both enzymes cleave factor V at the same location. The 25 amino acids determined by Edman degradation of the light chain of bovine factor V_a are identical to the sequence of the human protein except for threonine at position 3 and glutamic acid at position 21 (23). Although activation of human factor V with thrombin results in the cleavage of at least three bonds and removal of the large connecting fragment(s), activation with Russell's viper venom results in only one cleavage, which generates the light chain fragment. Factor V activated with Russell's viper venom has a specific activity that is identical to that of factor V activated by thrombin. In contrast, Eaton *et al.* (48) have

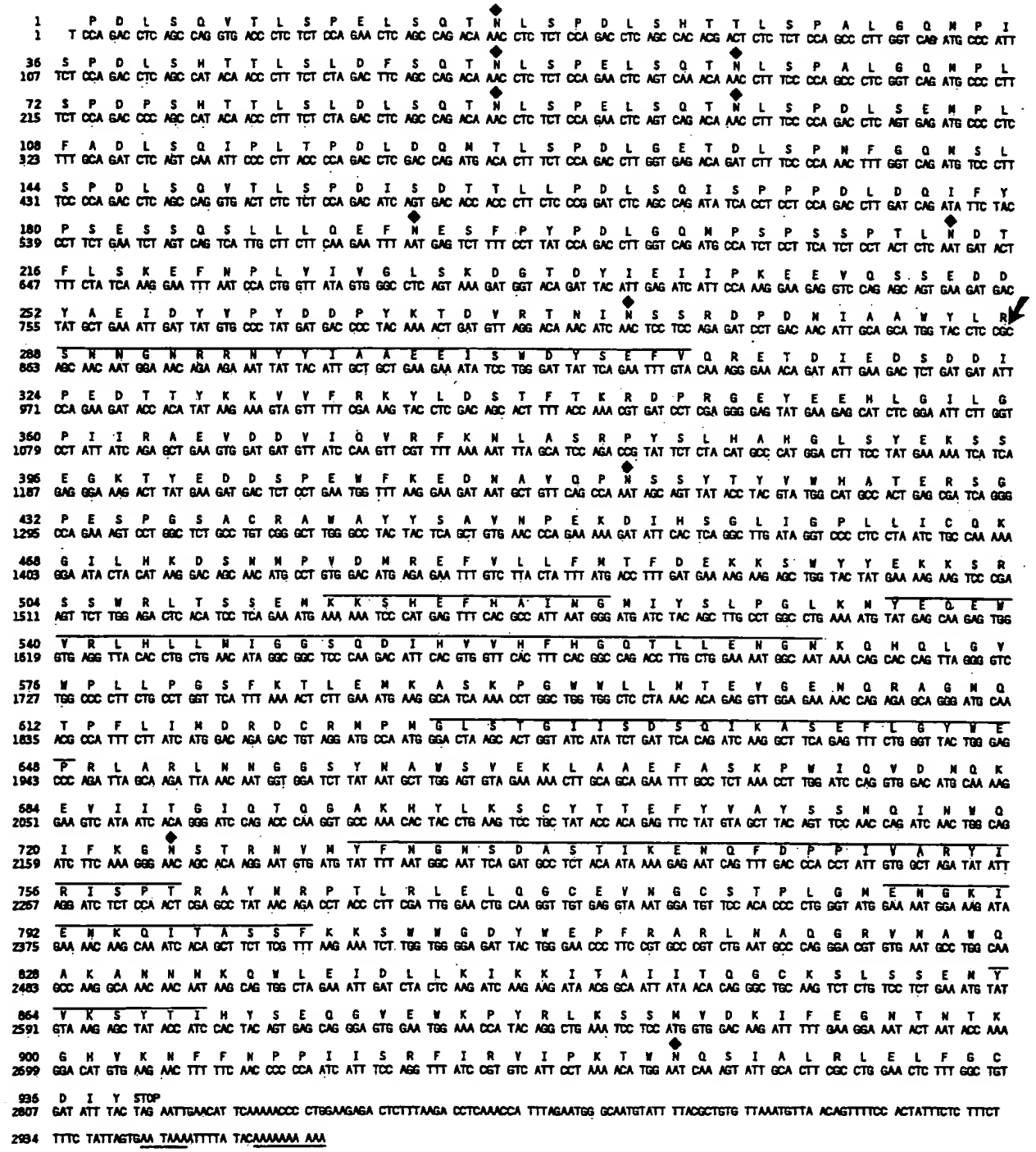

FIG. 2. Nucleotide sequence of the cDNA insert in λHV2970 that codes for the carboxyl terminus of human factor V. Numbering for the predicted amino acid sequence (in single-letter code) is arbitrary and will require revision when the complete amino acid sequence of human factor V is determined. The amino acid residues identified by Edman degradation are overlined. Potential N-linked carbohydrate attachment sites are shown by solid diamonds. A thrombin cleavage site is indicated by a heavy arrow, and the polyadenylylation or processing sequence and the poly(A) tail are underlined.

proposed that cleavage of three peptide bonds is required for complete activation of human factor VIII by thrombin.

The most striking feature of the carboxyl terminus of the connecting fragment of factor V_a is that it contains at least 20 repeats of a sequence of nine amino acids (Fig. 4). The DNA sequence coding for these repeats is strongly conserved and codes for a consensus sequence of Thr-Leu-Ser-Pro-Asp-Leu-Ser-Gln-$\frac{Thr}{Met}$. The carboxyl terminus of the connecting fragment (287 residues) is very acidic, with a calculated net charge of −37 at pH 7.0 (excluding carbohydrate). It is also rich in proline (12.2%), serine (15.7%), threonine (8.0%), and leucine (15.2%). It is apparently heavily glycosylated because there are eight potential N-linked oligosaccharide attachment sites within the 287 residues thus far identified. This is consistent with the fact that the connecting fragment(s) does not stain with Coomassie blue and migrates anomalously in

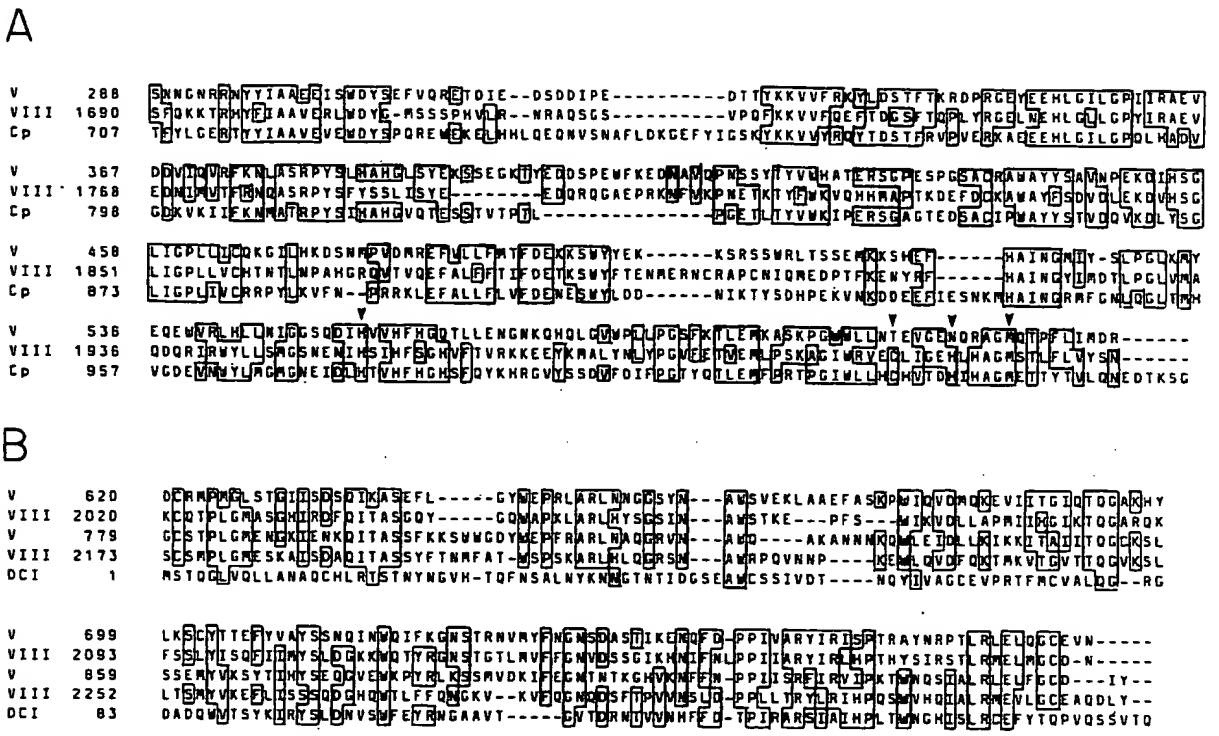

FIG. 3. (*A*) Amino acid sequence (in single-letter code) homology between the A domains in factor V, factor VIII, and ceruloplasmin. Sequences from factor V (residues 288–619), factor VIII (residues 1690–2332), and ceruloplasmin (residues 707–1046) were aligned by using the ALIGN computer program of Dayhoff *et al.* (43). Conserved amino acids are enclosed in boxes. The arrows designate the location of the proposed ligands for type I copper in ceruloplasmin. (*B*) Amino acid sequence homology between the C domains of factor V, factor VIII, and discoidin I. The sequences from factor V (residues 620–778 and 779–938), factor VIII (residues 2020–2172 and 2173–2332), and discoidin I (residues 1–160) were aligned by using the ALIGN computer program of Dayhoff *et al.* (43). Residues conserved in three or more sequences are enclosed in boxes.

$NaDodSO_4$/gel systems (4, 49). The function of this unusual region remains obscure. A search of the Protein Identification Resource* failed to identify any other proteins that are homologous with the 20 repeats of nine amino acids that are present in factor V. Furthermore, the carboxyl terminus of the connecting fragment does not show any significant amino acid sequence homology with the corresponding carbohydrate-rich connecting region of human factor VIII.

In the present investigation, ≈40% of the amino acid sequence of factor V has been established. These findings confirm that factor V, factor VIII, and ceruloplasmin comprise a family of related proteins that have arisen through gene duplication of the A domain (23, 24). Additional cDNA clones must be obtained to determine the domain structure of the remaining 60% of the factor V molecule and the structural features that give rise to its cofactor activity.

*National Biomedical Research Foundation (1986) Protein Sequence Database of the Protein Identification Resource (Natl. Biomed. Res. Found., Washington, DC 20007), Release No. 8.0.

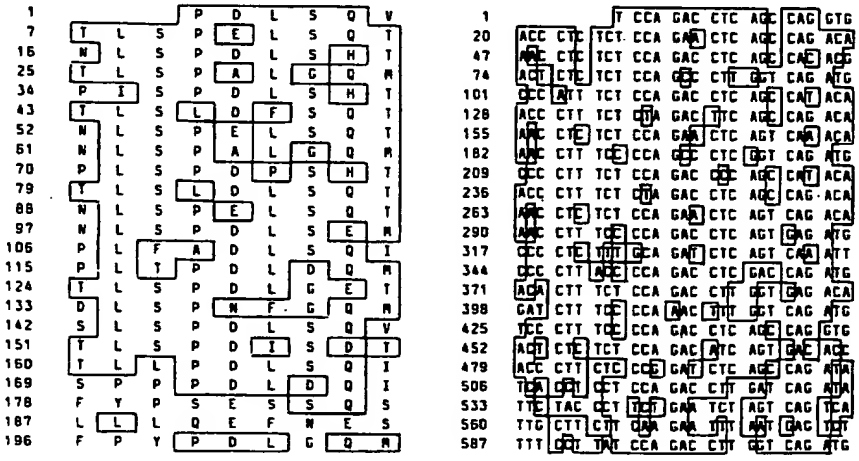

FIG. 4. Tandem repeats in the amino acid sequence (in single-letter code) and the cDNA sequence for the connecting region of human factor V. Amino acid residues or nucleotides that are identical with a consensus sequence of TLSPDLSQT/M or ACCCTTTCTCCAGACCTCAG-TCAGACA are enclosed in boxes.

The authors thank Drs. Donald Foster, Dominic Chung, Kazuo Fujikawa, Akitada Ichinose, Benito Que, Steven Leytus, Barbara Schach, and Kotoku Kurachi for helpful discussions and assistance. We also thank Brad McMullen for assistance in the amino acid sequence analysis, Dr. Fred Hagen for providing the Hep G2 λgt11 cDNA expression library, Dr. Frans Peetoom for kindly making available human plasma for the preparation of factor V, and Mrs. Lois Swenson for her assistance in the preparation of the manuscript. This work was supported by a research grant (HL 16919) from the National Institutes of Health.

1. Owren, P. A. (1947) *Acta Med. Scand. Suppl.* **194**, 1–316.
2. Nesheim, M. E., Myrmel, K. H., Hibbard, L. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 508–517.
3. Esmon, C. T. (1979) *J. Biol. Chem.* **254**, 964–973.
4. Kane, W. H. & Majerus, P. W. (1981) *J. Biol. Chem.* **256**, 1002–1007.
5. Katzmann, J. A., Nesheim, M. E., Hibbard, L. S. & Mann, K. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 162–166.
6. Suzuki, K., Dahlback, B. & Stenflo, J. (1982) *J. Biol. Chem.* **257**, 6556–6564.
7. Freeman, J. P., Guillen, M. S., Bezeaud, A. & Jackson, C. M. (1977) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **36**, 675 (abstr.).
8. Esmon, C. T., Owen, W. G., Duiguid, O. L. & Jackson, C. M. (1973) *Biochim. Biophys. Acta* **310**, 289–294.
9. Kane, W. H. & Majerus, P. W. (1982) *J. Biol. Chem.* **257**, 3963–3969.
10. Tracy, P. B., Peterson, J. M., Nesheim, M. E., McDuffie, F. C. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 10354–10361.
11. Tracy, P. B. & Mann, K. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2380–2384.
12. Maruyama, I., Salem, H. H. & Majerus, P. W. (1984) *J. Clin. Invest.* **74**, 224–230.
13. Bloom, J. W., Nesheim, M. E. & Mann, K. G. (1979) *Biochemistry* **20**, 4419–4425.
14. van de Waart, P., Bruls, H., Hemker, H. C. & Lindhout, T. (1983) *Biochemistry* **22**, 2427–2432.
15. Higgins, D. V. & Mann, K. G. (1983) *J. Biol. Chem.* **258**, 6503–6508.
16. Pusey, M. L. & Nelsestuen, G. L. (1984) *Biochemistry* **23**, 6202–6210.
17. Nesheim, M. E., Taswell, J. B. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 10952–10962.
18. Kisiel, W., Canfield, W. M., Ericsson, L. H. & Davie, E. W. (1977) *Biochemistry* **16**, 5824–5831.
19. Walker, F. J., Sexton, P. W. & Esmon, C. T. (1979) *Biochim. Biophys. Acta* **571**, 333–342.
20. Salem, H. H., Broze, G. J., Miletich, J. P. & Majerus, P. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1584–1588.
21. Salem, H. H., Esmon, N. L., Esmon, C. T. & Majerus, P. W. (1984) *J. Clin. Invest.* **73**, 968–972.
22. Salem, H. H., Broze, G. J., Miletich, J. P. & Majerus, P. W. (1983) *J. Biol. Chem.* **258**, 8531–8534.
23. Fass, D. N., Hewick, R. M., Knutson, G. J., Nesheim, M. E. & Mann, K. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1688–1691.
24. Church, W. R., Jernigan, R. L., Toole, J., Hewick, R. M., Knopf, J., Knutson, G. J., Nesheim, M. E., Mann, K. G. & Fass, D. N. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6934–6937.
25. Mann, K. G., Lawler, C. M., Vehar, G. A. & Church, W. R. (1984) *J. Biol. Chem.* **259**, 12949–12951.
26. Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkings, R. N., Tuddenham, E. G. D., Lawn, R. M. & Capon, D. J. (1984) *Nature (London)* **312**, 337–342.
27. Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, B., Coe, M. L., Knutson, G. J., Fass, D. N. & Hewick, R. M. (1984) *Nature (London)* **312**, 342–347.
28. Takahashi, N., Ortel, T. L. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 390–394.
29. Wilson, D. B., Salem, H. H., Mruk, J. S., Maruyama, I. & Majerus, P. W. (1984) *J. Clin. Invest.* **73**, 654–658.
30. Cerveny, T. J., Fass, D. N. & Mann, K. G. (1984) *Blood* **63**, 1467–1474.
31. Chiu, H. C., Schick, P. K. & Colman, R. W. (1985) *J. Clin. Invest.* **75**, 339–346.
32. Canfield, W. M. & Kisiel, W. (1982) *J. Clin. Invest.* **70**, 1260–1272.
33. Hagen, F. S., Gray, C. L., O'Hara, P., Grant, F. J., Saari, G. C., Woodbury, R. G., Hart, C. E., Insley, M., Kisiel, W., Kurachi, K. & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2412–2416.
34. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
35. Maxam, A. W. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
36. Silhavy, T. J., Berman, W. L. & Enquist, L. W. (1986) *Experiments with Gene Fusions* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 140–141.
37. Degen, S. J. F., MacGillivray, R. T. A. & Davie, E. W. (1983) *Biochemistry* **22**, 2087–2097.
38. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
39. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & Kurachi, K. (1985) *Biochemistry* **24**, 3736–3750.
40. Gross, R. H. (1986) *Nucleic Acids Res.* **14**, 591–596.
41. McMullen, B. A. & Fujikawa, K. (1985) *J. Biol. Chem.* **260**, 5328–5341.
42. Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H. & Davie, E. W. (1972) *Biochemistry* **11**, 4899–4903.
43. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
44. Ryden, L. & Bjork, I. (1976) *Biochemistry* **15**, 3411–3417.
45. Ryden, L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6767–6771.
46. Poole, S., Firtel, R. A., Lamar, E. & Rowekamp, W. (1981) *J. Mol. Biol.* **153**, 273–289.
47. Gabius, H. J., Springer, W. R. & Barondes, S. H. (1985) *Cell* **42**, 449–456.
48. Eaton, D., Rodriguez, H. & Vehar, G. A. (1986) *Biochemistry* **25**, 505–512.
49. Tal, M., Silberstein, A. & Nusser, E. (1985) *J. Biol. Chem.* **260**, 9976–9980.

Evidence Appendix E

Biochemistry 1992, 31, 3777-3785

Express Mail
Mailing Label No. EV832482014US

Structure of the Gene for Human Coagulation Factor V[†,‡]

Larry D. Cripe,[§] Karen D. Moore, and William H. Kane*,[∥]

Division of Hematology-Oncology, Departments of Medicine and Pathology, Box 3656 Duke University Medical Center, Durham, North Carolina 27710

Received December 13, 1991; Revised Manuscript Received February 5, 1992

ABSTRACT: Activated factor V (Va) serves as an essential protein cofactor for the conversion of prothrombin to thrombin by factor Xa. Analysis of the factor V cDNA indicates that the protein contains several types of internal repeats with the following domain structure: A1–A2–B–A3–C1–C2. In this report we describe the isolation and characterization of genomic DNA coding for human factor V. The factor V gene contains 25 exons which range in size from 72 to 2820 bp. The structure of the gene for factor V is similar to the previously characterized gene for factor VIII. Based on the aligned amino acid sequences of the two proteins, 21 of the 24 intron–exon boundaries in the factor V gene occur at the same location as in the factor VIII gene. In both genes, the junctions of the A1–A2 and A2–A3 domains are each encoded by a single exon. In contrast, the boundaries between domains A3–C1 and C1–C2 occur at intron–exon boundaries, which is consistent with evolution through domain duplication and exon shuffling. The connecting region or B domain of factor V is encoded by a single large exon of 2820 bp. The corresponding exon of the factor VIII gene contains 3106 bp. The 5′ and 3′ ends of both of these exons encode sequences homologous to the carboxyl-terminal end of domain A2 and the amino-terminal end of domain A3 in ceruloplasmin. There is otherwise no homology between the B domain exons. These data provide further insight into the evolutionary relationships within this family of related plasma proteins and provide a basis from which to begin the investigation of the cellular regulation of factor V biosynthesis and characterization of molecular defects in congenital factor V deficiency.

The generation of thrombin by the prothrombinase complex is a critical event in hemostasis and thrombosis. This complex consists of factor Xa, factor Va, prothrombin, calcium, and a phospholipid or cellular surface. Factor Va participates as an essential protein cofactor in the activation of prothrombin by factor Xa (Kane & Davie, 1988). Most of the factor V in whole blood circulates in the plasma; however, 25% of the factor V in blood is stored in platelet α-granules (Tracy et al., 1982). Platelet factor V is released during platelet activation and appears to be important for platelet surface prothrombin activation and normal hemostasis (Miletich et al., 1978). The major sites of factor V biosynthesis appear to include both the liver (Wilson et al., 1984) and megakaryocytes (Gewirtz et al., 1986). We (Kane & Davie, 1986; Kane et al., 1987) and others (Jenny et al., 1987) have previously determined the predicted amino acid sequence of human factor V by cDNA cloning. The mature protein consists of 2196 amino acids. Analysis of the factor V cDNA indicates that the protein contains several types of internal repeats organized with the following domain structure: A1–A2–B–A3–C1–C2. The A, B, and C domains each contain approximately 350, 836, and 150 amino acids, respectively. The primary sequence of human factor V is 40% identical to human coagulation factor VIII except in the B domain where there is no homology (Toole et al., 1984; Vehar et al., 1984). The role of factor VIIIa in the activation of factor X is analogous to the role of factor Va in prothrombin activation. Thus, the factor X activation complex consists of factor IXa, factor VIIIa, factor X, calcium, and a phospholipid or cellular surface. The A domains present in factor V and factor VIII are approximately 30% identical to the triplicated A domains present in the plasma copper-binding protein ceruloplasmin (Ortel et al., 1984). The C domains in factor V and factor VIII are also approximately 40% identical to the duplicated C domains present in a recently characterized murine breast epithelial cell protein (Stubbs et al., 1990).

The amino acid sequence homology among this family of proteins suggests that the genes for these proteins have evolved through a process of gene duplication and exon shuffling (Gilbert, 1985). The gene for factor VIII has been characterized (Gitschier et al., 1984), and it spans >180 kb on human chromosome Xq28. One striking feature of the factor VIII gene is that the entire B domain or connecting region is encoded by a particularly large exon of 3106 bp. The gene for human factor V has been localized to human chromosome 1q21-25 (Wang et al., 1988). Recent gene mapping studies (Watson et al., 1990) indicate that the gene for human factor V is located within a 300-kb region that also includes the genes for the selectin family of leukocyte adhesion molecules. We now report the cloning and characterization of the gene for human factor V. Our data provide evidence that gene duplication and exon shuffling occurred during the evolution of this family of related plasma proteins and provide a basis from which to begin the investigation of the cellular regulation of factor V biosynthesis and characterization of molecular defects in congenital factor V deficiency (parahemophilia).

Materials and Methods

Materials. All restriction enzymes and DNA-modifying enzymes were purchased from Bethesda Research Laboratories or United States Biochemicals. ^{32}P-Labeled nucleotides and

[†] This work was supported by Grants HL43106 and DK01965 from the National Institutes of Health, Searle Scholar Award 89-G-130, and Established Investigator Award 91-181 from the American Heart Association and Smith-Kline Beecham Inc., with funds contributed in part by the North Carolina Affiliate of the American Heart Association.

[‡] The genetic sequence for factor V given in this paper has been submitted to GenBank under Accession Number J05368.

[§] Supported by NIH Program Physician Scientist Award.

[∥] AHA Established Investigator; 1989 Searle Scholar.

[α-^{35}S]dATPαS were obtained from Amersham. The Sequenase Version 2.0 DNA sequencing kit was obtained from United States Biochemicals. DNA amplification kits were obtained from Perkin-Elmer Cetus.

Isolation of Genomic Clones for Factor V. A human lymphocyte genomic library in λEMBL 3 (Clontech) and a human lung fibroblast genomic library in λFIX (Stratagene) were screened by the in situ plaque hybridization method of Benton and Davis (1977). A full-length factor V cDNA probe (Kane et al., 1990) was used to screen approximately 1×10^6 recombinant phage from the λFIX library. Approximately 8×10^5 phage from the λEMBL3 library were screened in a similar fashion using a restriction fragment (*Eco*RI–*Acc*I) corresponding to nucleotides 1–269 of the factor V cDNA. The cDNA probes were labeled with [^{32}P]dCTP to a specific activity of $>10^8$ cpm/μg in the presence of random hexanucleotide primers and DNA polymerase I (Feinberg & Vogelstein, 1983). Positive genomic clones were plaque purified using standard methods (Sambrook et al., 1989).

Southern Blot Analysis of Genomic Clones and Human Genomic DNA. Phage DNA from the genomic clones and high molecular weight human genomic DNA from human peripheral blood leukocytes were isolated using standard methods (Sambrook et al., 1989). The DNA was digested with the restriction enzymes *Sst*I and/or *Xba*I, separated on 0.7% agarose gels, and transferred to nitrocellulose membranes (Southern, 1975). The membranes were then hybridized with ^{32}P-labeled factor V cDNA probes or oligonucleotides in order to determine the location of exons and to verify that no rearrangements occurred during the construction of the phage library. The filters were washed for 30 min at moderate (0.2× SSC,[1] 65 °C) or high (0.2× SSC, 70 °C) stringency, and autoradiographs were exposed using XAR-5 film with intensifying screens (Du Pont). Restriction fragments isolated from *Sst*I or *Xba*I digests of the genomics clones were subcloned into pBluescript (Stratagene). These plasmid subclones were then mapped by restriction enzyme digestion and Southern blotting. These data were used to construct a genomic restriction map for the enzymes *Sst*I and *Xba*I. The gene map was confirmed by characterization of at least two independent clones for the gene and by Southern blot analysis of human genomic DNA.

DNA Sequence Analysis of the Isolated Genomic Clones. DNA sequences were determined by the dideoxy chain termination method (Sanger et al., 1977) using either universal or synthetic oligonucleotide primers. Single-stranded templates were prepared using M13mp18/19 or pBluescript (Sambrook et al., 1989). Plasmids employed for double-stranded sequencing were purified by cesium chloride banding (Sambrook et al., 1989). Approximately 95% of the DNA sequence for exons and exon–intron boundaries were sequenced on both strands. Sequences that deviated from previously published cDNA sequences were verified on independent templates on both strands. DNA sequences were stored and analyzed using a Macintosh cx computer and MacVector version 3.5 software (IBI). In most cases, exons were aligned to the genomic *Xba*I and *Sst*I map using restriction sites present in the exon or exon–intron boundary sequence. The locations of exons 15–18 were determined by PCR (Saiki et al., 1988) using the following pairs of oligonucleotide primers: 5′-ATACCTACGTATGGCATG-3′ (exon 15) and 5′-GAGTCTCCAAGAACTTCG-3′ (exon 16); 5′-ATTATCAGAAGAGCAAGG-3′ (intron 16) and 5′-CAGATTGCCTTTTCCCTG-3′ (intron 17); 5′-GGGATGATCTACAGCTTG-3′ (exon 17) and 5′-TAGTAATAGGCACTCTCC-3′ (intron 18). Typically, reactions contained 100 ng of a bacteriophage clone template, 1 μM each oligonucleotide, and 2.5 units of *Taq*I polymerase. The DNA was amplified for 40 cycles of denaturation (1 min at 94 °C), annealing (2 min at 45–60 °C), and extension (2 min at 72 °C) using a Coy thermal cycler. The magnesium chloride concentration was varied between 1 and 6 mM in order to optimize results. Reaction products were analyzed on 1% agarose gels with ethidium bromide staining.

Results and Discussion

Isolation and Characterization of Factor V Genomic Clones. Forty-five genomic phage clones were isolated from the λFIX library using a ^{32}P-labeled full-length factor V cDNA probe. These clones contained inserts ranging from 12 to 30 kb. Preliminary analysis indicated that 24 of these isolates represented unique overlapping clones. Southern blotting and partial DNA sequence analysis indicated that the sequences 5′ to nucleotide 327 in the factor V cDNA were not present in these clones. Subsequently, five clones were isolated from the λEMBL3 library by hybridization with a probe containing nucleotides 1–269 of the factor V cDNA. Restriction and Southern analysis of these clones demonstrated that three were unique overlapping clones with inserts ranging from 12 to 16 kb. Partial DNA sequence analysis of these clones indicated that they contained the sequences corresponding to nucleotides 1–326 of the factor V cDNA. Hybridization experiments demonstrated no overlap between the two sets of overlapping clones (see below).

Structure of the Human Factor V Gene. Twenty of the independent genomic clones were chosen for detailed characterization. Restriction fragments from the eight clones shown in Figure 1 were also subcloned into pBluescript and characterized as described under Materials and Methods. Characterization of *Sst*I and or *Xba*I digests of these clones or plasmid subclones by Southern blotting provided the data used to construct a restriction map of the factor V gene for these enzymes. It is possible that some small genomic fragments (<250 bp) containing intron sequences may not have been detected in our analysis. Restriction mapping of the genomic clones indicated that the gene was comprised of at least 17 *Sst*I fragments ranging in size from 0.7 to 12 kb (Figure 1). Fourteen of these fragments contained exons as judged by Southern blotting experiments. No genomic clones completely spanned the second intron. The sizes of the cloned genomic *Sst*I fragments were consistent with the size of the fragments observed when human genomic DNA was probed with factor V cDNA probes. This indicates that no major rearrangements occurred during the construction of the genomic libraries. When genomic DNA was probed with factor V cDNA probes encoding the heavy chain (nt 1–2074) or the B domain (nt 2268–4785), only the predicted hybridizing fragments were observed consistent with a single copy of the factor V gene. However when genomic DNA was probed with probes encoding the light-chain region (nt 5001–6903, 6000–6590, or 6591–6903), several additional hybridizing bands were observed after high stringency washes.[2] These additional bands may represent genomic DNA that is sequence related to the factor V gene.

The structure of the human factor V gene is shown in Figure 1. The human factor V gene spans >80 kb of DNA. The

[1] Abbreviations: SSC, 0.01 M sodium citrate buffer, pH 7.4, 0.15 M sodium chloride; PCR, polymerase chain reaction.

[2] L. D. Cripe and W. H. Kane, unpublished observations.

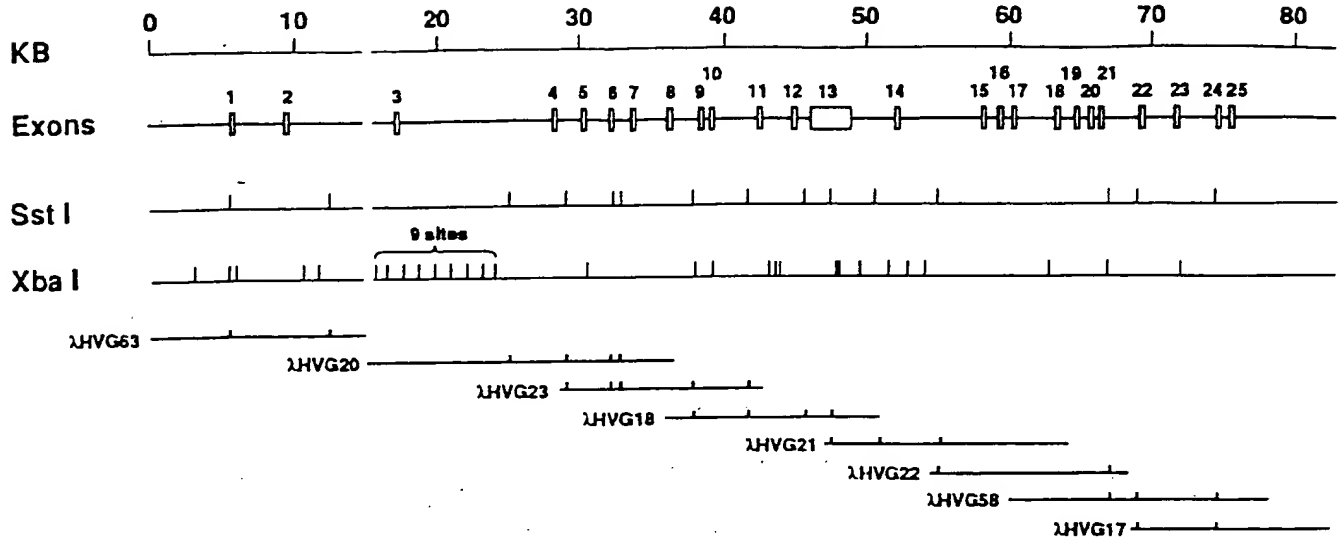

FIGURE 1: Organization of the human coagulation factor V gene. The size scale in kilobases (kb) is shown at the top of the figure. The second line is schematic representation of the human factor V gene drawn to scale with exons indicated by the rectangles connected by lines representing the introns. The number for each exon is indicated. Genomic restriction sites for *Sst*I and *Xba*I are depicted below. The order of the eight *Xba*I fragments that are bracketed was not determined. The location for eight of the unique overlapping human genomic clones is indicated by the solid lines at the bottom of the figure. The identity of each clone is shown at the right of the bar, and the vertical lines indicate the location of genomic *Sst*I sites contained in each clone. Clone λHVG63 was obtained from the λEMBL 3 library; the remainder of the clones were isolated from the λFIX library. There is no overlap between clones λHVG63 and λHVG20. This gap occurs in intron 2 of the factor V gene and it is indicated in the map by the interrupted lines.

nucleotide sequences of exons and exon–intron boundaries were sequenced by the dideoxy chain termination method. The gene consists of 25 exons which encode the factor V mRNA as determined by comparison with the cloned cDNA (Kane & Davie, 1986; Kane et al., 1987; Jenny et al., 1987). The exons range in size from 72 to 2820 bp (Table I). The genomic clones include ~5.4 kb of DNA 5′ to exon 1 and ~7 kb 3′ to exon 25. The introns in the factor V gene range in size from 0.4 to ≥11 kb (Table II). There is no cross-hybridization between clones λHVG63 and λHVG20 indicating that these clones are missing a portion of intron 2. Splice donor and acceptor sequences agree with the GT-AG rule and conform to the consensus sequence of Mount (1982) except for intron 6. Intron 6 agrees with the consensus sequence except that the donor–acceptor sequence is GC-AT. This variant donor splice site sequence accounts for the majority of nonconforming donor splice sites in nonimmunoglobulin genes (Shapiro & Senepathy, 1987).

Comparison of Factor V Genomic and cDNA Sequences. The sequence of the factor V genomic clones is in excellent agreement with the two previously published cDNAs (Kane & Davie, 1986; Kane et al., 1987; Jenny et al., 1987). There are seven positions with differences between the genomic sequence and the two published cDNAs (Table III). In two cases the differences occur in domain A1 at nucleotides 481 and 628. These differences are silent in that they do not affect the predicted amino acid sequence; however, our previously reported cDNA (Kane et al., 1987) contains a *Pst*I site at position 477 which is not present in the genomic clones or the cDNA sequence reported by Jenny (Jenny et al., 1987). The five remaining differences occur within the B domain of factor V. Four of these differences result in amino acid substitutions.

The human factor V cDNAs isolated by Kane (Kane et al., 1987) and Jenny (Jenny et al., 1987) include 5′ untranslated regions of 76 and 90 nucleotides, respectively. The first 20 nucleotides of these sequences are not identical. Comparison of the cDNA sequences to the genomic sequence helps to explain these differences. The sequence reported by Jenny (Jenny et al., 1987) includes the sequence GAATTCCG which is not present in the gene (Figure 2). This sequence corresponds to the *Eco*RI linker used to construct the cDNA library. The sequence of the Jenny cDNA following the linker agrees completely with the genomic sequence. The first 11 nucleotides of the cDNA sequence that we reported previously do not agree with the genomic sequence reported here. This discrepancy appears to be an artifact that occurred near adjacent *Alu*I sites during the construction of the cDNA library. Thus, nucleotides 1–16 of our cDNA correspond to the comple-

Table I: Exon Organization of the Human Factor V Gene[a]

exon	cDNA	length	amino acids	domain
1	14–234	221	53	5′ untranslated sequence, signal sequence, A1
2	235–326	90	30	A1
3	327–449	124	41	A1
4	450–662	212	71	A1
5	663–806	144	48	A1
6	807–1028	221	74	A1
7	1029–1194	166	56	A1-A2
8	1195–1372	178	59	A2
9	1373–1472	99	33	A2
10	1473–1687	215	72	A2
11	1688–1838	151	50	A2
12	1839–2051	213	71	A2
13	2052–4872	2820	940	A2-A3, connecting region or B domain, 18 and 9 amino acid tandem repeats, thrombin cleavage sites
14	4873–5047	175	59	A3
15	5048–5284	237	79	A3
16	5285–5495	211	70	A3
17	5496–5675	180	60	A3
18	5676–5792	117	39	A3
19	5793–5864	72	24	C1
20	5865–5968	104	35	C1
21	5969–6124	156	52	C1
22	6125–6269	145	48	C1
23	6270–6421	152	51	C2
24	6422–6604	182	61	C2
25	6605–6890	286	48	C2, 3′ untranslated sequence

[a] Numbering of nucleotide residues and protein domains corresponds to the previously published sequences (Kane & Davie, 1988). Amino acids with interrupted codons were assigned to the exon containing two of the three codon nucleotides.

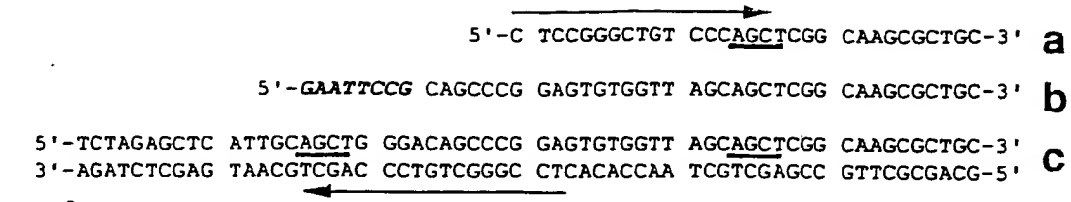

FIGURE 2: Comparison of the factor V genomic sequence with 5′ untranslated regions of published cDNA sequences. The sequences corresponding to the cDNA reported by (a) Kane (Kane et al., 1987), (b) Jenny (Jenny et al., 1987), and (c) the genomic sequence reported here are listed. Sequences corresponding to the *Eco*RI linker are shown in bold print. *Alu*I restriction sites are underlined. The first 18 nucleotides of the cDNA reported by Kane (Kane et al., 1987) differ from the sequence reported by Jenny (Jenny et al., 1987) but are identical to complementary strand genomic sequence (arrows).

Table II: Intron–Exon Splice Junctions of the Human Factor V Gene[a]

Intron	Exon	Intron	Exon	Intron size (kb)	Intron Type
1	TCAAG	GTAACTCA.....TTATTTCCAG	TTTGAATC	3.8	II
2	TTCAG	GTAAGAAC.....GTTTTCACAG	GACTTCTT	>10	I
3	AGAAG	GTAAGATA.....ACTCCTCCAG	GTGCTTCT	11	I
4	AAAAG	GTAAGAAC.....CCTTTCATAG	GGACCCTA	2.2	I
5	GCCAG	GTAACACG.....CTTTTCTCAG	ATATAACA	1.9	I
6	GCAAG	GCAAGAAA.....CCTAACTCAG	CTGGGATG	1.4	I
7	GACAA	GTGAGTTG.....TCACTTTAAG	AAAATACA	2.4	II
8	TCAAA	GTAGTAAC.....ATTACTTGAG	ATCGTGTT	1.9	0
9	CTCAG	GTTTGAAT.....TTTCTTTCAG	GCAGGAAC	0.6	I
10	TACAG	GTACTTTT.....TGTCCTCCAG	AGGGCAGC	3.1	0
11	GAGCA	GTAAGTCA.....TCACTTTCAG	CTATCAAT	2.2	I
12	TGTTG	GTGAGTAA.....TCTTCTGTAG	GAACTTGG	1.1	I
13	CAAAG	GTTTGGCC.....ATTATTTCAG	GGAAACAG	2.9	II
14	TCCAA	GTAAGTCT.....CTTTTCCTAG	GTTCGTTT	5.8	0
15	ACCCA	GTACTTCT.....TGTTTTCCAG	GAAAAAGA	1.2	0
16	TCACG	GTATTTTT.....TTGTCTTCAG	CCATTAAT	1.0	I
17	GCCTG	GTAAAGAT.....TCTATTTTAG	GTTCATTT	3.0	I
18	CAGAG	GTATCACA.....TTTTTCATAG	ACTGTAGG	1.3	I
19	TCTGG	GTAAGTTG.....CTTGAGTTAG	GTTACTGG	1.1	I
20	TCCAG	GTTTGTCT.....ATATTGGCAG	GTGGACAT	0.4	0
21	TGATG	GTTTGTGT.....TTTCTTCTAG	TATTTTAA	2.7	0
22	AAATG	GTAAGGTA.....CTTTTGATAG	GATGTTCC	2.3	I
23	CCAAG	GTCAAGTA.....ATTTTCATAG	GCAAACAA	2.8	0
24	ACAAG	GTAGAGTG.....TTTTGGTCAG	ATTTTTGA	1.0	0
Consensus					
	C	A	TTTTTT T		
	AG	GT AGT.....	N AG		
	A	G	CCCCCC C		

[a] Intron numbers are indicated in the first column. The nucleotide sequences of the exon–intron boundaries are shown for each intron. The splice donor and acceptor dinucleotides are in boldface type and match the consensus GT...AG except for intron 6. The sizes of introns in the human factor V gene were determined by restriction enzyme digestion or PCR. Type 0 indicates a splice between codons, type I and type II indicate a splice after the first or second nucleotide of a codon, respectively. The consensus sequence proposed by Mount (1982) is shown at the bottom of the table.

Table III: Differences in Nucleotide and Predicted Amino Acid Sequences among Factor V cDNA and Genomic Clones[a]

nucleotide no.	genomic codon	Jenny et al. cDNA codon	Kane et al. cDNA codon	amino acid
481	CGG	GCG	GCA	Ala
628	TCG	TCG	TCT	Ser
2377	TCA	TCG	TCG	Ser
2649	AAA	AGA	AGA	Lys to Arg
2670	CAT	CGT	CGT	His to Arg
2849	AAG	GAG	GAG	Lys to Glu
3930	CTT	CTT	ATT	Leu to Ile

[a] The first column indicates the nucleotide number corresponding to the cDNA sequences reported by Kane (Kane & Davie, 1986; Kane et al., 1987). The second column indicates the sequence of the codon determined from the factor V genomic clones in the present work. The nucleotide that differs from previously reported nucleotides is shown in bold print. The second and third columns show the codons reported by Jenny (Jenny et al., 1987) and Kane (Kane & Davie, 1986; Kane et al., 1987). The fifth column indicates the amino acid encoded by the genomic sequence followed by the amino acid encoded by a cDNA sequence if different.

mentary strand of the genomic sequence terminating in an upstream *Alu*I site (Figure 2). RNAse protection assays with a probe that extends 147 bases upstream from the ATG initiation codon demonstrate two protected fragments of 97 and 103 nucleotides.[3] Primer extension assays will be used to verify these as authentic transcription start sites. Thus, the first exon of the factor V gene encodes a 5′ untranslated region of 97 or 103 nucleotides, a 28 amino acid leader peptide, and the first 24 amino acids of the mature protein.

Exon–Intron Organization of the Human Factor V Gene. Comparison with the Factor VIII Gene and Correlation with Protein Domains. The exon–intron organization of the gene for human factor V is almost identical to the organization of the gene for human factor VIII (Gitschier et al., 1984). The genes for factors V and VIII are composed of 25 and 26 exons, respectively. The larger mRNA of factor VIII relative to factor V is primarily due to the 1.8 kb of 3′ untranslated

[3] L. D. Cripe and W. H. Kane, unpublished observations.

sequence in factor VIII. The gene for factor V spans ≥80 kb whereas the gene for factor VIII is considerably larger at ~180 kb. This difference is largely due to the fact that six of the introns in the factor V gene are much smaller than the corresponding introns in the factor VIII gene. We have previously used the computer programs of Dayhoff to align the amino acid sequences for factor V and factor VIII (Kane & Davie, 1988). Figure 3 shows the aligned amino acid sequences near the exon–intron boundaries. Inspection of the corresponding genomic DNA sequences for factor V and factor VIII reveals that in 21 of 24 cases the exon–intron boundaries occur at precisely the same location in the aligned sequence. The 3′ boundary of exon 5 of the factor V gene corresponds to the 3′ boundary of exon 6 in the gene for factor VIII. This can be explained either by the gain or the loss of an intron subsequent to the divergence of these two genes. The 3′ exon–intron boundaries for exon 9 and 13 and the 5′ exon–intron boundary for exon 14 are not precisely conserved as judged by the protein alignment. These discrepancies may either be due to inaccuracies in the protein alignment or alternatively to the phenomenon of intron drift (Craik et al., 1983).

Although the complete structure of the gene for ceruloplasm has not been reported, a pseudogene for ceruloplasmin on chromosome 8 has been characterized as well as several clones for the functional ceruloplasmin gene on chromosome 3 (Koschinsky et al., 1987; Royle et al., 1987). Two exon–intron boundaries for the ceruloplasmin gene were reported. One boundary, corresponding to the 3′ region of exon 9 in the factor V gene, does not occur at the same position as the boundary in either the factor V or factor VIII gene on the basis of the protein alignments. The second boundary corresponds to the 5′ end of exon 13 in the factor V gene and exon 14 in the factor VIII gene. This boundary occurs at precisely the same location in all three genes on the basis of the protein alignments.

Figure 4 shows the location of introns and protein domains for factor V and factor VIII. Within the individual A domains of both proteins, the locations of some introns are conserved suggesting that they existed in the primordial gene for this domain whereas the others were either gained or lost subsequent to the duplication events. The sequences encoding the A1–A2 and A2–A3 domain boundaries are each contained within a single exon in both genes. There is amino acid homology between factor V, factor VIII, and ceruloplasmin in the sequences encoded by the 5′ and 3′ ends of these exons; however, there are major differences in the amino acid sequences encoded by the central portion of each exon (Ortel et al., 1984; Jenny et al., 1987; Kane et al., 1987). Thus, exon 8 of the factor VIII gene contains a 33 amino acid insert that includes 15 acidic residues which is not present in exon 7 of the factor V gene. Futhermore, the exons encoding the domain A2–A3 boundaries in the genes for factor V and factor VIII are both extremely large and encode the entire B domain for the protein. Exon 13 in the factor V gene spans 2820 bp whereas the exon 14 in the factor VIII gene is 3106 bp in length. The exons encoding the B domain in factor V and factor VIII are much larger in size than the reported average size for exons in vertebrate genes of 133 bp (Smith, 1988). Recently, however, large exons ranging from 1.5 to 8.7 kb have been described in a number of genes or gene families including apolipoprotein B (Ludwig et al., 1987), the large subunit of RNA polymerase II (Ahearn et al., 1987), nuclear hormone receptor genes (Faber et al., 1989), collagen genes (Muragaki et al., 1991), and mucin genes (Ligtenberg et al., 1990; Toribara et al., 1991).

The B domain of factor V is not homologous to any known protein; however, it shares several features with the family of mucin glycoproteins (Ligtenberg et al., 1990; Toribara et al., 1991; Gum et al., 1989; Porchet et al., 1991). These similarities may be an example of convergent evolution. Thus, the genes for mucins and factor V contain large exons which encode amino acid sequences that are rich in hydroxyl amino acids and proline. These proteins contain tandem repeats of short amino acid sequences and appear to be heavily glycosylated with both O- and N-linked oligosaccharide chains. In each of these proteins the amino acid sequences and tandem repeat structures appear to be unique. The genes for mucins have been shown to be polymorphic, containing a variable number of tandem repeats (Ligtenberg et al., 1990; Toribara et al., 1991; Gum et al., 1989; Porchet et al., 1991). Thus far, this kind of polymorphism has not been observed in the gene for human factor V.[4] Factor V contains 31 repeats of a nine amino acid sequence. One of the cDNA clones that was isolated from a HepG2 cDNA library contained an in-frame deletion of precisely 11 of the tandem nine amino acid repeats (Kane & Davie, 1986; Kane et al., 1987). It will be of interest to determine whether this was an artifact of the cDNA library construction or a polymorphism in the human factor V gene. On the other hand, the predicted amino acid sequence for bovine factor V contains eight fewer nine amino acid tandem repeats and only a single copy of the 17 amino acid sequence that is duplicated in the human protein (Guinto et al., 1989). The B domain of factor VIII is also rich in carbohydrate and hydroxyl amino acids; however, it does not contain any tandem repeat structures (Toole et al., 1984; Vehar et al., 1984).

In contrast to the sequences encoding the A1–A2 and A2–A3 domain boundaries, the sequences encoding the A3–C1 and C1–C2 domain boundaries are not each contained within a single exon. In these cases the domain boundaries in the protein correlate precisely with type I exon–intron boundaries in the gene. Recently, Stubbs et al. (1990) cloned the cDNA for a murine breast epithelial cell protein of unknown function that consists of a single epidermal growth factor like domain followed by the two C-type domains. The C domains in the epithelial cell protein are approximately 40% identical to the C domains in factor V and factor VIII. The structure of the gene encoding this protein is not known. On the basis of the amino acid sequence homology, we predict that the exon–intron organization of this gene will be very similar to the genes for factor V and factor VIII.

Evolutionary and Functional Considerations. The function of the large connecting region or B domains in factor V and factor VIII remains poorly understood. Both proteins circulate in plasma as relatively inactive cofactors. Activation of both molecules by thrombin yields an active species comprising a calcium-dependent heterodimer composed of heavy and light chains. The heavy chains contain domains A1–A2, whereas the light chains contain domains A3–C1–C2. The B domains contain all three of the thrombin cleavage sites in factor V and two of the three thrombin cleavage sites in factor VIII. Following thrombin activation, the B domains of both proteins are released as very large activation peptides comprising >40% of the mass of the original molecule. Recent ultrastructural data obtained using electron microscopy suggests that the B domains of factor V and factor VIII form a rodlike tail extending from a globular domain containing the heavy chain and light chain (Mosesson et al., 1990a,b). The B domains of factor V and factor VIII are not required for generation

[4] L. D. Cripe and W. H. Kane, unpublished observations.

```
                                        .....EXON 1.....GAG CCC ACA AAC TCA AG GTAACTCA    V
                                                        Glu Pro Thr Asn Ser Se
                                                        Leu Pro Val Asp Ala Ar
                                        .....EXON 1.....CTG CCT GTG GAC GCA AG GTAAAGGC    VIII

          TTATTTCCAG  T TTG AAT CTT TCT GTA.....EXON 2.....CAA TCT ACC ATT TCA G  GTAAGAAC    V
                      r Leu Asn Leu Ser Val                Gln Ser Thr Ile Ser G
                      g Phe Pro Pro Arg Val                Arg Pro Pro Trp Met G
          CTTCACGCAG  A TTT CCT CCT AGA GTG.....EXON 2.....AGG CCA CCC TGG ATG G  GTAATGAA    VIII

          GTTTTCACAG GA CTT CTT GGG CCT ACT.....EXON 3.....AGT AAA TTA TCA GAA G  GTAAGATA    V
                      ly Leu Leu Gly Pro Thr                Ser Lys Leu Ser Glu G
                      ly Leu Leu Gly Pro Thr                Trp Lys Ala Ser Glu G
          ATTCTTACAG GT CTG CTA GGT CCT ACC.....EXON 3.....TGG AAA GCT TCT GAG G  GTGAGTAA    VIII

          ACTCCTCCAG GT GCT TCT TAC CTT GAC.....EXON 4.....CTT ATC TGT AAA AAA G  GTAAGAAC    V
                      ly Ala Ser Tyr Leu Asp                Leu Ile Cys Lys Lys G
                      ly Ala Glu Tyr Asp Asp                Leu Val Cys Arg Glu G
          CCTGCTATAG GA GCT GAA TAT GAT GAT.....EXON 4.....CTA GTA TGT AGA GAA G  GTAAGTGT    VIII

          CCTTTCATAG GG ACC CTA ACT GAG GGT.....EXON 5.....AAT GGG ACA ATG CCA G  GTAACACG    V
                      ly Thr Leu Thr Glu Gly                Asn Gly Thr Met Pro A
                      ly Ser Leu Ala Lys Glu                Asn Arg Ser Leu Pro G
          TCTTTTTCAG GG AGT CTG GCC AAG GAA....EXON 5-6....AAC AGG TCT CTG CCA G  GTATGTAC    VIII

          CTTTTCTCAG AT ATA ACA GTT TGT GCC.....EXON 6.....CCA AAA CAT TTG CAA G  GCAAGAAA    V
                      sp Ile Thr Val Cys Ala                Pro Lys His Leu Gln A
                      ly Leu Ile Gly Cys His                Ser Ser His Gln His A
          TACTTTACAG GT CTG ATT GGA TGC CAC.....EXON 7.....TCT TCC CAC CAA CAT G  GTAATATC    VIII

          CCTAACTCAG CT GGG ATG CAG GCT TAC.....EXON 7.....CCA GCG AAT ATG GAC AA GTGAGTTG    V
                      la Gly Met Gln Ala Tyr                Pro Ala Asn Met Asp Ly
                      sp Gly Met Glu Ala Tyr                Leu Ala Pro Asp Asp Ar
          CTGACTCCAG AT GGC ATG GAA GCT TAT.....EXON 8.....CTC GCC CCC GAT GAC AG GTAAGCAC    VIII

          TCACTTTAAG  A AAA TAC AGG TCT CAG.....EXON 8.....AGA GAC ACA CTC AAA    GTAGTAAC    V
                      s Lys Tyr Arg Ser Gln                Arg Asp Thr Leu Lys
                      g Ser Tyr Lys Ser Gln                Gly Asp Thr Leu Leu
          TCTTATACAG  A AGT TAT AAA AGT CAA.....EXON 9.....GGA GAC ACA CTG TTG    GTAAGTTG    VIII

          ATTACTTGAG    ATC GTG TTC AAA AAT.....EXON 9.....TCT TCT TTC ACC TCA G  GTTTGAAT    V
                        Ile Val Phe Lys Asn                Ser Ser Phe Thr Ser G
                        Ile Ile Phe Lys Asn                Leu Pro Lys G
          ATACTTTCAG    ATT ATA TTT AAG AAT.....EXON 10....TTA CCA AAA G          GTAAATAT    VIII

          TTTCTTTCAG GC AGG AAC AAC ACC ATG.....EXON 10....AGG CGA GGA ATA CAG    GTACTTTT    V
                      ly Arg Asn Asn Thr Met                Arg Arg Gly Ile Gln
                      ly Val Lys His Leu Lys                Gln Arg Gly Asn Gln
          TTGTGGGTAG GT GTA AAA CAT TTG AAG.....EXON 11....CAA AGA GGA AAC CAG    GTGAGTTC    VIII

          TGTCCTCCAG    AGG GCA GCA GAC ATC.....EXON 11....TCA AAC ATC ATG AGC A  GTAAGTCA    V
                        Arg Ala Ala Asp Ile                Ser Asn Ile Met Ser T
                        Ile Met Ser Asp Lys                Ser Asn Ile Met His S
          TAATTAACAG    ATA ATG TCA GAC AAG.....EXON 12....TCC AAC ATC ATG CAC A  GTGAGTAA    VIII

          TCACTTTCAG CT ATC AAT GGC TAT GTG.....EXON 12....ACA ATG GAT AAT GTT G  GTGAGTAA    V
                      hr Ile Asn Gly Tyr Val                Thr Met Asp Asn Val G
                      er Ile Asn Gly Tyr Val                Ser Met Glu Asn Pro G
          GTTTTTGCAG GC ATC AAT GGC TAT GTT.....EXON 13....TCG ATG GAA AAC CCA G  GTTAGTTA    VIII

          ATTATTTCAG GA ACT TGG ATG TTA ACT.....EXON 13....CAA AG                 GTTTGGCC    V
                      ly Thr Trp Met Leu Thr                Gln Ar
                      ly Leu Trp Ile Leu Gly                Ser Ser Pro His Val Leu Arg Asn Ar
          TCATCTCCAG GT CTA TGG ATT CTG GGG.....EXON 14....AGC TCC CCA CAT GTT CTA AGA AAC AG GTATGAAT   VIII

TTAATTTCAG  G GAA ACA GAT ATT GAA GAC TCT GAT.....EXON 14....GAT GAT GTT ATC CAA    GTAAGTCT    V
            g Glu Thr Asp Ile Glu Asp Ser Asp                Asp Asp Val Ile Gln
                                    g Ala                    Glu Asp Asn Ile Met
            TTTTCTCCAG                G GCT.....EXON 15....GAA GAT AAT ATC ATG    GTGAGTTA    VIII


          CTTTTCCTAG    GTT CGT TTT AAA AAT.....EXON 15....TCA GCT GTG AAC CCA    GTACTTCT    V
                        Val Arg Phe Lys Asn                Ser Ala Val Asn Pro
                        Val Thr Phe Arg Asn                Ser Asp Val Asp Leu
          TGTTCTACAG    GTA ACT TTC AGA AAT.....EXON 16....TCT GAT GTT GAC CTG    GTGAGTTA    VIII

          TGTTTTCCAG    GAA AAA GAT ATT CAC.....EXON 16....TCC CAT GAG TTT CAC G  GTATTTTT    V
                        Glu Lys Asp Ile His                Ser His Glu Phe His A
                        Glu Lys Asp Val His                Asn Tyr Arg Phe His A
          CCCTCCCTAG    GAA AAA GAT GTG CAC.....EXON 17....AAT TAT CGC TTC CAT G  GTAATATA    VIII

          TTGTCTTCAG CC ATT AAT GGG ATG ATC.....EXON 17....TGG CCC CTT CTG CCT G  GTAAAGAT    V
                      la Ile Asn Gly Met Ile                Trp Pro Leu Leu Pro G
                      la Ile Asn Gly Tyr Ile                Tyr Asn Leu Tyr Pro G
          CCTTCTCCAG CA ATC AAT GGC TAC ATA.....EXON 18....TAC AAT CTC TAT CCA G  GTATGAGC    VIII
```

```
TCTATTTTAG GT TCA TTT AAA ACT CTT.....EXON 18....CTT ATC ATG GAC AGA G  GTATCACA    V
           ly Ser Phe Lys Thr Leu                 Leu Ile Met Asp Arg A
           ly Val Phe Glu Thr Val                 Leu Val Tyr Ser Asn L
TTTTTATAAG GT GTT TTT GAG ACA GTG.....EXON 19....CTG GTG TAC AGC AAT A  GTGAGTAG    VIII

TTTTTCATAG AC TGT AGG ATG CCA ATG.....EXON 19....GCT TCA GAG TTT CTG G  GTAAGTTG    V
           sp Cys Arg Met Pro Met                 Ala Ser Glu Phe Leu G
           ys Cys Gln Thr Pro Leu                 Ala Ser Gly Gln Tyr G
TTCATTTCAG AG TGT CAG ACT CCC CTG.....EXON 20....GCT TCA GGA CAA TAT G  GTAAATAC    VIII

CTTGAGTTAG GT TAC TGG GAG CCC AGA.....EXON 20....AAA CCT TGG ATC CAG    GTTTGTCT    V
           ly Tyr Trp Glu Pro Arg                 Lys Pro Trp Ile Gln
           ly Gln Trp Ala Pro Lys                 Phe Ser Trp Ile Lys
TTGGGCAAAG GA CAG TGG GCC CCA AAG.....EXON 21....TTT TCT TGG ATC AAG    GTTAGAAA    VIII

ATATTGGCAG    GTG GAC ATG CAA AAG.....EXON 21....ACA AGG AAT GTG ATG    GTTTGTGT    V
              Val Asp Met Gln Lys                 Thr Arg Asn Val Met
              Val Asp Leu Leu Ala                 Thr Gly Thr Leu Met
TAATTGGTAG    GTG GAT CTG TTG GCA.....EXON 22....ACT GGA ACC TTA ATG    GTATGTAA    VIII

TTTCTTCTAG    TAT TTT AAT GGC AAT.....EXON 22....GGT TGT GAG GTA AAT G  GTAAGGTA    V
              Tyr Phe Asn Gly Asn                 Gly Cys Glu Val Asn G
              Val Phe Phe Gly Asn                 Gly Cys Asp Leu Asn S
CTCCATACAG    GTC TTC TTT GGC AAT.....EXON 23....GGC TGT GAT TTA AAT A  GTAAGTGC    VIII

CTTTTGATAG GA TGT TCC ACA CCC CTG.....EXON 23....GCC TGG CAA GCC AAG    GTCAAGTA    V
           ly Cys Ser Thr Pro Leu                 Ala Trp Gln Ala Lys
           er Cys Ser Met Pro Leu                 Ala Trp Arg Pro Gln
TTTCTTTGAG GT TGC AGC ATG CCA TTG.....EXON 24....GCC TGG AGA CCT CAG    GTAAGAGG    VIII

ATTTTCATAG    GCA AAC AAC AAT AAG.....EXON 24....TCC ATG GTG GAC AAG    GTAGAGTG    V
              Ala Asn Asn Asn Lys                 Ser Met Val Asp Lys
              Val Asn Asn Pro Lys                 Asn Gly Lys Val Lys
TTGCCCTCAG    GTG AAT AAT CCA AAA.....EXON 25....AAT GGC AAA GTA AAG    GTAAGCTG    VIII

TTTTGGTCAG    ATT TTT GAA GGA AAT.....EXON 25....
              Ile Phe Glu Gly Asn
              Val Phe Gln Gly Asn
CCTCTTTCAG    GTT TTT CAG GGA AAT.....EXON 26....
```

FIGURE 3: Comparison of exon–intron junction sequences of the genes for factors V and VIII. The amino acid sequences of human factor V and factor VIII were aligned by the method of Dayhoff as previously reported (Kane & Davie, 1988). The first line of each comparison is the genomic sequence for factor V determined in the present work. The second line is the corresponding translated amino acid sequence for factor V. The third line is the translated amino acid sequence for factor VIII which has been aligned to the factor V amino acid sequence. The fourth line in each series is the corresponding sequence for the factor VIII gene determined by Gitschier (Gitschier et al., 1984). Genomic exon sequences are shown as triplets, and the translated amino acid sequence for split codons is indicated by the split in the three-letter amino acid code. The number for each exon in the factor V and the factor VIII genes is indicated. Exon 5 in the factor V gene contains the sequences corresponding to exons 5 and 6 in the factor VIII gene.

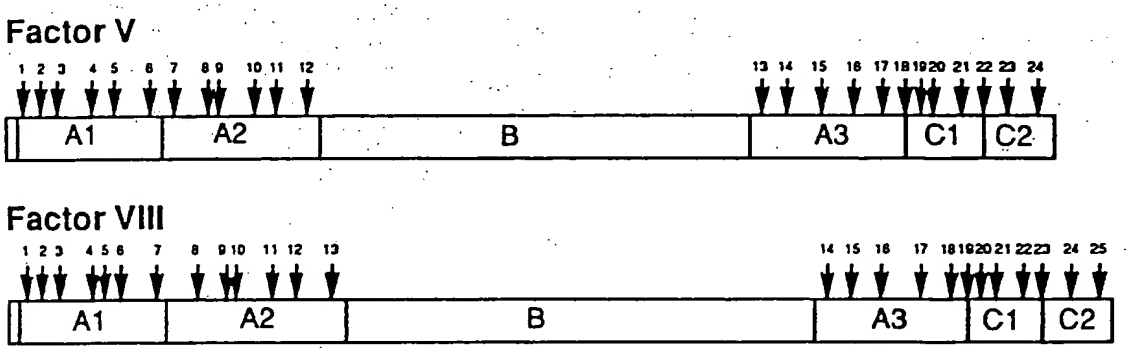

FIGURE 4: Comparison of the domain structures and exon–intron structures for factor V and factor VIII. The boxes represent the domain structures for factor V and factor VIII and are drawn to scale. The identities of the domains are indicated by the letters and correspond to the amino acid sequences described previously (Kane & Davie, 1988). The arrows indicate the location of introns which are removed from the mature mRNA. The number of each intron in the gene for factor V or factor VIII is indicated.

of procoagulant activity since deletion of a large portion of these regions in recombinant proteins does not reduce the procoagulant activity of either molecule (Toole et al., 1986; Kane et al., 1990). However, these observations do not exclude other functions for the B domains. Thus, the factor V deletion mutant appeared to have increased constitutive procoagulant activity suggesting that the entire B domain is necessary to inhibit cofactor activity (Kane et al., 1990). Complete deletion of the B domain results in a molecule with markedly diminished activity secondary to reduced rates of proteolytic activation by thrombin (Pittman et al., 1990). Furthermore, the B domain of factor VIII serves a unique function in that it contains the binding site for von Willebrand factor (Pittman & Kaufman, 1989). The von Willebrand factor–factor VIII interaction is required for the stabilization of factor VIII in vitro as well as in vivo (Kaufman et al., 1988).

Because of the large size of the exons encoding the B domains in factor V and factor VIII, we speculate that these exons arose through the insertion of a processed gene derived from reverse transcription of an mRNA. It is possible that the exons encoding the B domains of both factor V and factor VIII are derived from a single primordial cofactor gene. In this case, the lack of amino acid homology between the two cofactors would be due to divergent evolution. Analysis of the amino acid sequences for factor V and factor VIII from different species suggests that there has been relatively little

selective pressure to conserve amino acid sequences in the B domains. Thus, the predicted amino acid sequence for the B domain of bovine factor V is only 59% identical to the corresponding sequence of the human protein whereas there is 88% amino acid sequence identity in the heavy chain and light chain regions (Guinto et al., 1989). Similar divergence has been observed between human and porcine factor VIII sequences. Toole et al. (1986) reported ~50% amino acid identity in the B domain region whereas there was 80–85% amino acid identity in the heavy-chain and light-chain regions. Divergence within the B domain of human factor V is further suggested by the fact that the five discrepancies between reported genomic and cDNA sequences that result in missense mutations all occur within the B domain. While the available data are consistent with a single insertion event in the primordial cofactor gene followed by divergent evolution, it is also possible that two separate insertion events occurred subsequent to the divergence of the genes for factor V and factor VIII. An insertion event also appears to have occurred in the A1–A2 exon in the factor VIII gene which codes for an additional 33 amino acids that are not present in the corresponding region of factor V or ceruloplasmin. Characterization of related genes in primitive species may further clarify the evolution of this family of related genes.

Characterization of the gene for human factor V provides the groundwork for examining natural mutations that affect factor V function and result in the rare bleeding disorder congenital factor V deficiency. Functional characterization of these defects should provide insight into the regulation of the prothrombinase complex. The major sites of synthesis of factor V appear to be the hepatocyte and megakaryocyte (Wilson et al., 1984; Gewirtz et al., 1986). The acquired factor V deficiency associated with hepatic failure and the correlation of clinical severity with platelet factor V levels in parahemophilia indicate the importance of factor V synthesis at these sites (Miletich et al., 1978). Factor V has also been reported to be synthesized in other cell types including monocytes (Altieri & Edgington, 1989), endothelial cells (Shen & Edgington, 1991), T-cells (Shen et al., 1990), and vascular smooth muscle cells (Rodgers, 1988). The importance of cellular surfaces in regulating blood coagulation suggests a possible role for factor V associated with these cells. Characterization of the regulatory elements of the factor V gene may help explain the tissue-specific expression of this protein and the significance of extrahepatic synthesis in physiological and pathophysiological states.

ACKNOWLEDGMENTS

We thank Mr. Thad Howard for his assistance with the genomic Southern blot experiments. We also thank the following individuals for helpful discussions and criticisms during the course of this work: Drs. G. R. Vandenbark, Michael F. Seldin, Rebecca Oakey, Thomas L. Ortel, and Charles S. Greenberg.

Registry No. V_2, 65522-14-7.

REFERENCES

Ahearn, J. M., Barolomei, M. S., West, M. L., Cisek, L. J., & Corden, J. L. (1987) *J. Biol. Chem. 262*, 10695.

Altieri, D. C., & Edgington, T. S. (1989) *J. Biol. Chem. 264*, 2969.

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180.

Craik, C. S., Rutter, W. J., & Fletterick, R. (1983) *Science 220*, 1125.

Faber, P. W., Kuiper, G. G. J. M., van Rooij, H. C. J., van der Korput, J. A. G. M., Brinkman, A. O., & Trapman, J. (1989) *Mol. Cell. Endocrinol. 61*, 257.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6.

Gewirtz, A. M., Keefer, M., Doshi, K., Annamalai, A. E., Chiu, H. C., & Colman, R. W. (1986) *Blood 67*, 1639.

Gilbert, W. (1985) *Science 226*, 823.

Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J., & Lawn, R. M. (1984) *Nature 312*, 326.

Guinto, E. R., Odegaard, B., Mann, K. G., & MacGillivray, R. T. A. (1989) *Thromb. Haemostasis 62*, 80 (Abstract).

Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A., & Kim, Y. S. (1989) *J. Biol. Chem. 264*, 6480.

Jenny, R. J., Pittman, D. D., Toole,J. J., Kriz, R. W., Aldape, R. A., Hewick, R. M., Kaufman, R. J., & Mann, K. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4846.

Kane, W. H., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6800.

Kane, W. H., & Davie, E. W. (1988) *Blood 71*, 539.

Kane, W. H., Ichinose, A., Hagen, F. S., & Davie, E. W. (1987) *Biochemistry 26*, 6508.

Kane, W. H., Devore-Carter, D., & Ortel, T. L. (1990) *Biochemistry 29*, 6762.

Kaufman, R. J., Wasley, L. C., & Dorner, A. J. (1988) *J. Biol. Chem. 263*, 6352.

Koschinsky, M. L., Chow, B. K., Schwartz, J., Hamerton, J. L., & MacGillivray, R. T. (1987) *Biochemistry 26*, 7760.

Ligtenberg, M. J. L., Vos, H. L., Gennissen, A. M. C., & Hilkens, J. (1990) *J. Biol. Chem. 265*, 5573.

Ludwig, E. W., Blackhart, B. D., Pierotti, V. I., Caiati, L., Fortier, C., Knott, T., Scott, J., Mahley, R. W., Levy-Wilson, B., & McCarthy, B. J. (1987) *DNA 6*, 363.

Miletich, J. P., Majerus, D. W., & Majerus, P. W. (1978) *J. Clin. Invest. 62*, 824.

Mosesson, M. W., Church, W. R., DiOrio, J. P., Krishnaswamy, S., Mann, K. G., Hainfeld, J. F., & Wall, J. S. (1990a) *J. Biol. Chem. 265*, 8863.

Mosesson, M. W., Fass, D. N., Lollar, P., DiOrio, J. P., Parker, C. G., Knutson, G. J., Hainfeld, J. F., & Wall, J. S. (1990b) *J. Clin. Invest. 85*, 1983.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459.

Muragaki, Y., Jacenko, O., Apte, S., Mattei, M., Ninomiya, Y., & Olsen, B. R. (1991) *J. Biol. Chem. 266*, 7721.

Ortel, T. L., Takahashi, N., & Putnamm, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4761.

Pittman, D. D., & Kaufman, R. J. (1989) *Thromb. Haemostasis 61*, 161.

Pittman, D. D., Thompkinson, K., Marquette, K., Jenny, R. J., Mann, K. B., & Kaufman, R. J. (1990) *Blood 76* (Suppl.) 433a (Abstract).

Porchet, N. N., Van Cong, J., Dufosse, J. P., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A., & Aubert, J. P. (1991) *Biochem. Biophys. Res. Commun. 175*, 414.

Rodgers, G. M. (1988) *Biochim. Biophys. Acta 968*, 17.

Royle, N. J., Irwin, D. M., Koschinsky, M. L., MacGillivray, R. T., & Hamerton, J. L. (1987) *Somatic Cell Mol. Genet. 13*, 285.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487.

Sambrook, J. E., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
Shapiro, M. B., & Senepathy, P. (1987) *Nucleic Acids Res. 15*, 7155.
Shen, N. L. L., & Edgington, T. S. (1991) *Arterioscler. Thromb. 11*, 1543a (Abstract).
Shen, N. L. L., Altieri, D. C., & Edgington, T. S. (1990) *FASEB J. 4*, A2275 (Abstract).
Smith, M. W. (1988) *J. Mol. Evol. 27*, 45.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503.
Stubbs, J. D., Lekutis, C., Singer, K. L., Bui, A., Yuzuki, D., Srinivasan, U., & Parry, G. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8417.
Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, W. B., Coe, M. L., Knutson, G. J., Fass, D. N., & Hewick, R. M. (1984) *Nature 312*, 342.
Toole, J. J., Pittman, D. D., Orr,E. C., Murtha, P., Wasley, L. C., & Kaufman, R. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5939.
Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Petersen, G. M., & Kim, Y. S. (1991) *J. Clin. Invest. 88*, 1005.
Tracy, P. B., Eide, L. L., Bowie, E. J., & Mann, K. G. (1982) *Blood 60*, 59.
Vehar, G. A., Keyt, B., Eaton, D. L., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkins, R. N., Tuddenham, E. G. D., Lawn, R. M., & Capon, D. J. (1984) *Nature 312*, 337.
Wang, H., Riddell, D. C., Guinto, E. R., & MacGillivray, R. T. A. (1988) *Genomics 2*, 234.
Watson, M. L., Kingsmore, S. F., Johnston, G. I., Siegelman, M. H., Le Beau, M. M., Lemons, R. S., Bora, N. S., Howard, T. A., Weissman, I. L., McEver, R. P., & Seldin, M. F. (1990) *J. Exp. Med. 172*, 263.
Wilson, D. B., Salem, H. H., Mruk, J. S., Maruyama, I., & Majerus, P. W. (1984) *J. Clin. Invest. 73*, 654.

Identification of Amino Acids Modified by the Bifunctional Affinity Label 5′-(*p*-Fluorosulfonyl)benzoyl)-8-azidoadenosine in the Reduced Coenzyme Regulatory Site of Bovine Liver Glutamate Dehydrogenase†

Kenneth E. Dombrowski,‡ Yu-Chu Huang, and Roberta F. Colman*
Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716
Received December 24, 1991

ABSTRACT: Bovine liver glutamate dehydrogenase reacts with the bifunctional affinity label 5′-(*p*-(fluorosulfonyl)benzoyl)-8-azidoadenosine (5′-FSBAzA) in a two-step process: a dark reaction yielding about 0.5 mol of -SBAzA/mol of subunit by reaction through the fluorosulfonyl moiety, followed by photoactivation of the azido group whereby covalently bound -SBAzA becomes cross-linked to the enzyme [Dombrowski, K. E., & Colman, R. F. (1989) *Arch. Biochem. Biophys. 275*, 302–308]. We now report that the rate constant for the dark reaction is not reduced by ADP or GTP, but it is decreased 7-fold by 2 mM NADH and 40-fold by 2 mM NADH + 0.2 mM GTP, suggesting that 5′-FSBAzA reacts at the GTP-dependent NADH inhibitory site. The amino acid residues modified in each phase of the reaction have been identified. Modified enzyme was isolated after each reaction phase, carboxymethylated, and digested with trypsin, chymotrypsin, or thermolysin. The digests were fractionated by chromatography on a phenylboronate agarose column followed by HPLC. Gas-phase sequencing of the labeled peptides identified Tyr^{190} as the major amino acid which reacts with the fluorosulfonyl group; Lys^{143} was also modified but to a lesser extent. The predominant cross-link formed during photolysis is between modified Tyr^{190} and the peptide Leu^{475}-Asp^{476}-Leu^{477}-Arg^{478}, which is located near the C-terminus of the enzyme. Thus, 5′-FSBAzA is effective in identifying critical residues distant in the linear sequence, but close within the regulatory nucleotide site of glutamate dehydrogenase.

Bovine liver glutamate dehydrogenase [L-glutamate:NAD(P)$^+$ oxidoreductase (deaminating), EC 1.4.1.3] is an allosteric enzyme that is activated by ADP, but inhibited by GTP and high concentrations of NADH. The enzyme in its smallest, active form is a hexamer of six identical subunits (Goldin & Frieden, 1972; Julliard & Smith, 1979). For each subunit there are six binding sites for catalytic and regulatory purine nucleotides: one catalytic site that binds either NAD(H) or NAD(P)H (Goldin & Frieden, 1972), one additional coenzyme site which is regulatory (Krause et al., 1974), two ADP sites (Batra & Colman, 1986a), and two GTP binding sites as measured in the presence of NADH (one of high affinity and one of low affinity) (Pal & Colman, 1979).

Considerable information regarding the amino acid residues within the catalytic and regulatory sites of bovine liver glutatmate dehydrogenase has come from affinity labeling ex-

†This work was supported by USPHS Grant DK 37000 (to R.F.C.), NSF Grant DMB-9105116 (to R.F.C.), and NIH Postdoctoral Fellowship GM12936 (to K.E.D.).
‡Present address: Department of Internal Medicine, Texas Tech University Health Sciences Center, 1400 Wallace Blvd., Amarillo, TX 79124.

0006-2960/92/0431-3785$03.00/0

Evidence Appendix F

Express Mail
Mailing Label No. EV832482014US

MOLECULAR BIOLOGY OF THE CELL

Bruce Alberts • Dennis Bray
Julian Lewis • Martin Raff • Keith Roberts
James D. Watson

Garland Publishing, Inc.
New York & London

"Long ago it became evident that the key to every biological problem must finally be sought in the cell, for every living organism is, or at sometime has been, a cell."

Edmund B. Wilson
The Cell in Development and Heredity
3rd edition, 1925, Macmillan, Inc.

Bruce Alberts received his Ph.D. from Harvard University and is currently a Professor in the Department of Biochemistry and Biophysics at the University of California Medical School in San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Senior Scientist in the Medical Research Council Cell Biophysics Unit at King's College London. *Julian Lewis* received his D.Phil. from Oxford University and is currently a Lecturer in the Anatomy Department at King's College London. *Martin Raff* received his M.D. degree from McGill University and is currently a Professor in the Zoology Department at University College London. *Keith Roberts* received his Ph.D. from Cambridge University and is currently Head of the Department of Cell Biology at the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from the University of Indiana and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

Library of Congress Cataloging in Publication Data
Main entry under title:

Molecular biology of the cell.

Includes bibliographies and index.
1. Cytology. 2. Molecular biology. I. Alberts, Bruce, 1938– [DNLM: 1. Cells. 2. Molecular biology. QH 581.2 M718]
QH581.2.M64 1983 574.87 82-15692
ISBN 0-8240-7282-0

Published by Garland Publishing, Inc.
136 Madison Avenue, New York, NY 10016

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5

Recombinant DNA Technology[27]

In the early 1970s DNA was the most difficult cellular compound for the biochemist to analyze. Enormously long and chemically monotonous, the nucleotide sequence of the hereditary material could be approached only by indirect means—such as through protein or RNA sequencing or by genetic analysis. Today the situation has entirely changed. From being the hardest macromolecule of the cell to analyze, DNA has become the easiest. It is now possible to excise specific regions of DNA, to obtain them in essentially unlimited quantities, and to determine the sequence of their nucleotides at a rate of several hundred nucleotides a day.

The new **recombinant DNA technology** has provided powerful and novel approaches to understanding the complex mechanisms by which eucaryotic gene expression is regulated, and it has largely superceded conventional methods for determining the amino acid sequence of a protein. Elaborations of the same methods offer great commercial promise for the large-scale economical production of protein hormones and vaccines, available at present only with great labor and cost.

Recombinant DNA technology comprises a mixture of techniques, some new and some borrowed from other fields such as microbial genetics (Table 4–12). The most important ones are (1) specific cleavage of DNA by *restriction nucleases,* (2) *nucleic acid hybridization,* which makes it possible to identify specific sequences of DNA or RNA with great accuracy and sensitivity by their ability to bind a complementary nucleic acid sequence, (3) *DNA cloning,* whereby a specific DNA fragment is integrated into a rapidly replicating genetic element (plasmid or virus) so that it can be amplified in bacteria or yeast cells, and (4) *DNA sequencing* of the nucleotides in a cloned DNA fragment.

Restriction Nucleases Hydrolyze DNA Molecules at Specific Nucleotide Sequences[28]

Many bacteria make enzymes called **restriction nucleases,** which protect them by degrading any invading foreign DNA molecules. Each enzyme recognizes a specific sequence of four to six nucleotides in DNA. The correspond-

Table 4–12 Major Steps in the Development of Recombinant DNA Technology

1869	**Miescher** isolated DNA for the first time.
1944	**Avery** provided evidence that DNA, rather than protein, carries the genetic information during bacterial transformation.
1953	**Watson and Crick** proposed the double-helix model for DNA structure based on x-ray results of **Franklin and Wilkins.**
1961	**Marmur and Doty** discovered DNA renaturation, establishing the specificity and feasibility of nucleic acid hybridization reactions.
1962	**Arber** provided the first evidence for the existence of DNA restriction enzymes, leading to their later purification and use in DNA sequence characterization by **Nathans and H. Smith.**
1966	**Nirenberg, Ochoa,** and **Khorana** elucidated the genetic code.
1967	**Gellert** discovered DNA ligase, the enzyme used to join DNA fragments together.
1972–1973	DNA cloning techniques were developed by the laboratories of **Boyer, Cohen, Berg,** and their colleagues at Stanford University and the University of California at San Francisco.
1975–1977	**Sanger and Barrell** and **Maxam and Gilbert** developed rapid DNA-sequencing methods.

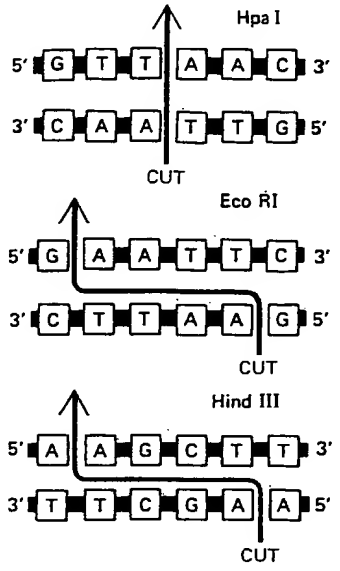

Figure 4–46 The DNA nucleotide sequences recognized by three widely used restriction nucleases. Such sequences are often, as in these examples, six base pairs long and "palindromic"—that is, the nucleotide sequences of the two strands are the same in the recognized region. The two strands of DNA are cut at or near the recognition sequence, often with a staggered cleavage that creates a cohesive end—as for Eco RI and Hind III. Restriction nucleases are obtained from various species of bacteria: Hpa I is from *Hemophilus parainfluenzae*; Eco RI, *Escherichia coli*; and Hind III, *Hemophilus influenzae*.

ing sequences in the genome of the bacterium itself are "camouflaged" by methylation at an A or a C residue, but any foreign DNA molecule that enters the cell is immediately recognized by the nuclease, and both strands of its DNA helix are cut (Figure 4–46). Many restriction nucleases have been purified from different species of bacteria, and more than 100, most of which recognize different nucleotide sequences, are now commercially available.

A particular restriction nuclease will cut any long length of DNA double helix into a series of fragments known as **restriction fragments.** By comparing the sizes of the DNA fragments produced from a particular genetic region after treatment with a combination of different restriction nucleases, a **restriction map** can be constructed that shows the location of each cutting (restriction) site in relation to its neighbors. Since such maps reflect the arrangement of selected nucleotide sequences in the region, a comparison of such maps for two or more related genes will give a rough estimate of the homology between them. For example, the restriction maps, and therefore presumably the nucleotide sequences, of the entire chromosomal regions coding for hemoglobin chains in man, orangutan, and chimpanzee have remained largely unchanged during the 5 to 10 million years since these species first diverged (Figure 4–47).

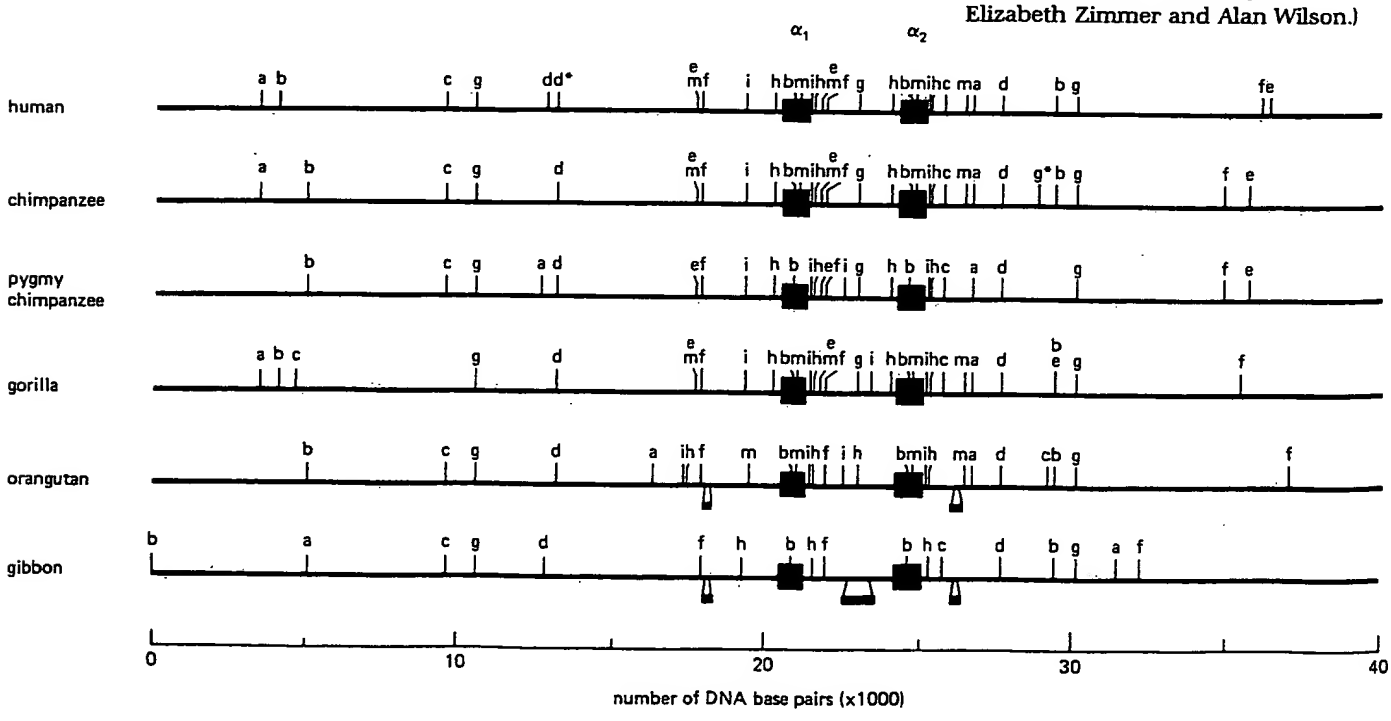

Figure 4–47 Restriction maps of human and various primate DNAs in a cluster of genes coding for hemoglobin. The two squares in each map indicate the positions of the DNA corresponding to the α-globin genes. Each letter stands for a site cut by a different restriction nuclease. The location of each cut was determined by comparing the sizes of the DNA fragments generated by treating the DNAs with the various restriction nucleases, individually and in combinations. (Courtesy of Elizabeth Zimmer and Alan Wilson.)

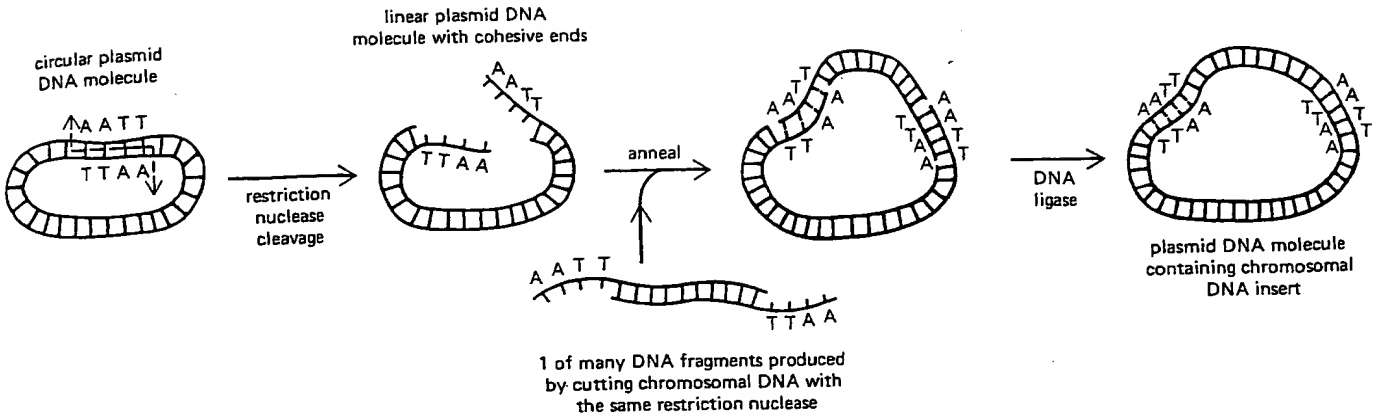

Figure 4–48 The cohesive ends produced by many kinds of restriction nucleases (see Figure 4–46) allow two DNA fragments to be joined by complementary nucleotide base-pair interactions. DNA fragments that are joined in this way can be covalently linked in a reaction catalyzed by the enzyme DNA ligase. In this example, a hybrid plasmid DNA molecule that contains a chromosomal DNA insert is formed.

Many restriction nucleases produce staggered cuts, which leave short, single-stranded ends on both fragments. These are known as *cohesive ends* since they can form complementary base pairs with any other end produced by the same enzyme. A circular DNA molecule that is cut at a single site by this type of restriction nuclease will therefore tend to re-form a circle by the annealing (base-pairing) of its cohesive ends. The cohesive ends generated by restriction enzymes have been very important in recombinant DNA technology because they enable any two DNA fragments to be joined, provided that they were generated with the same restriction nuclease, and thus have complementary cohesive ends. Once the two ends have joined by complementary base-pairing, they can be sealed by an enzyme known as a *DNA ligase*, which forms covalent phosphodiester bonds between the opposing ends of each strand of DNA (Figure 4–48). The combined use of restriction enzymes and DNA ligase has made it possible to graft fragments of any DNA into self-replicating elements.

Selected DNA Sequences Are Produced in Large Amounts by Cloning[29]

Fragments of DNA from any source can be amplified more than a millionfold by inserting them into a *plasmid* or a bacterial virus (*bacteriophage*) and then growing these in bacterial (or yeast) cells—a process called **DNA cloning.** Plasmids are small circular molecules of double-stranded DNA that occur naturally in both bacteria and yeast, where they replicate as independent units as the host cell proliferates. Although they generally account for only a small fraction of the total host cell DNA, they often carry vital genes, such as those that confer resistance to antibiotics. These genes, and the relatively small size of the plasmid DNA, are exploited in recombinant DNA technology.

Because it is so much smaller, plasmid DNA can easily be separated from the DNA of the host cell and purified. For use as *cloning vectors*, such purified plasmid DNA molecules are cut once with a restriction nuclease and then annealed to the DNA fragment that is to be cloned. The hybrid plasmid DNA molecules produced are then reintroduced into bacteria that have been made transiently permeable to macromolecules. Only some of the treated cells will take up a plasmid. They can be selected by the antibiotic resistance conferred on them by the plasmid since they alone will grow in the presence of antibiotic. As these bacteria divide, the plasmid also replicates to produce an enormous number of copies of the original DNA fragment (Figure 4–49). At the end of the period of proliferation, the hybrid plasmid DNA molecules are

purified and the copies of the original DNA fragments excised by a second treatment with the same restriction endonuclease (Figure 4–50).

The DNA to be cloned is often obtained by cleaving the entire genome of a cell with a specific restriction endonuclease. An enormous number of DNA fragments is obtained in this way—anywhere between 10^5 to 10^7 fragments from a mammalian genome, for example. The cloning process, therefore, may produce millions of different bacterial or yeast colonies, each harboring a plasmid with a different inserted genomic DNA sequence. The rare colony whose plasmid contains the genomic DNA region of interest must then be selected and allowed to proliferate to form a large cell population, or *clone*. The selection of the desired colony is often the most difficult part of the cloning procedure. The technique normally used for identifying the colony containing a specific cloned DNA fragment involves the use of radioactive nucleic acid probes complementary to the cloned DNA. We shall now discuss how such probes are commonly made.

Copies of Specific mRNA Molecules Can Be Cloned[30,31]

The cloning procedure just described is sometimes called a "shotgun" approach because the entire genomic DNA is cut into an enormous number of fragments that are randomly placed with respect to genes. As a result, some will contain parts of genes and many will contain only *noncoding* DNA and thus no genes at all. An alternative strategy is to begin the cloning process by selecting only those DNA sequences that are transcribed into RNA. This is done by extracting the mRNA (or a purified subfraction of the mRNA) from cells and then making a *DNA copy* (called a **cDNA molecule**) of each mRNA molecule present. This is made possible by an enzyme known as *reverse transcriptase* because, instead of catalyzing the transcription of DNA into RNA, it catalyzes the reverse process of synthesizing a complementary DNA chain on an RNA template. The single-stranded cDNA molecules synthesized by reverse transcriptase can be converted into double-stranded cDNA molecules (by using the enzyme *DNA polymerase*), inserted into plasmids, and cloned (Figure 4–51).

It is possible to construct plasmids in a way that allows the cloned cDNA to direct the synthesis within a cell of large amounts of the particular protein that the cDNA specifies. By means of such "genetic engineering," bacteria or yeast can be induced to make useful proteins, such as human insulin, growth hormone, and interferon, in enormous quantities.

Alternatively, cDNA can be used to identify the rare gene-containing clones produced by the shotgun approach. Here the procedure is to make single-stranded cDNA using radioactive nucleotide precursors. The resulting radiolabeled DNA can be hybridized to the complementary genomic clone in a way that will be described below. Because the cDNA is made from mRNA, it will

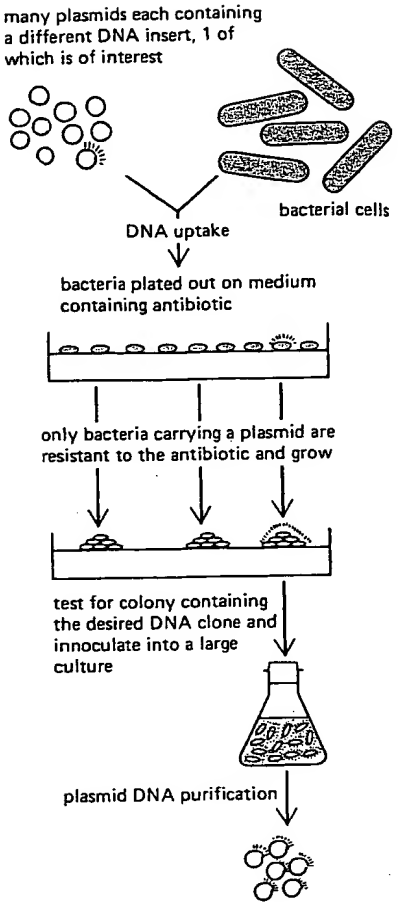

Figure 4–49 Purification and amplification of a specific DNA sequence by DNA cloning in a bacterium. DNA fragments are cloned in yeast cells by a similar procedure.

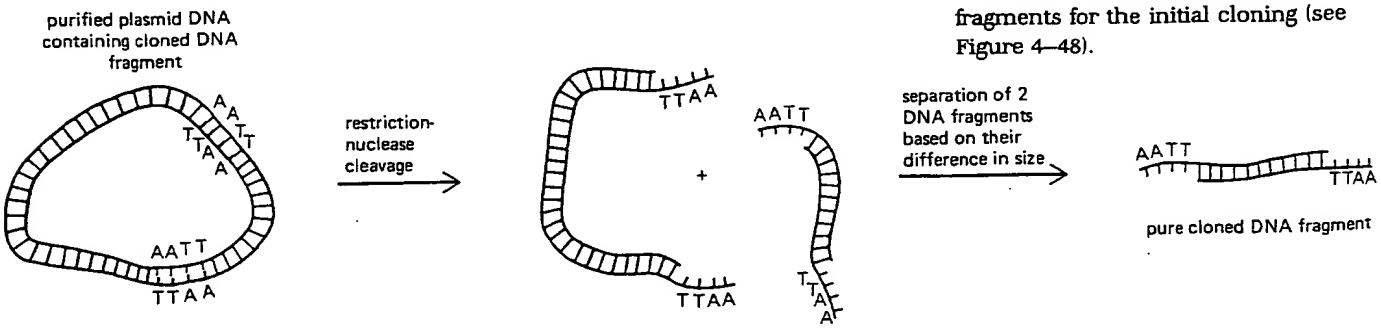

Figure 4–50 Recovery of a cloned DNA fragment from a plasmid containing a recombinant DNA molecule. The fragment is cut out of the plasmid by the same restriction nuclease that created the DNA fragments for the initial cloning (see Figure 4–48).

correspond to genomic DNA that codes for a protein, and its hybridization to the DNA in a clone marks that clone as one containing part of a gene encoding the mRNA molecule.

Cloned DNA Fragments Can Be Rapidly Sequenced[32]

It has recently become possible to determine the nucleotide sequence of cloned DNA fragments simply and quickly. The principle underlying one of these methods is illustrated in Figures 4–52 and 4–53. As a result of this new technology, the complete DNA sequences of more than 100 mammalian genes have already been determined, including those coding for insulin, hemoglobin, interferon, and cytochrome c. At present, the easiest and most accurate way to sequence the amino acids in a protein is by sequencing its gene and then using the genetic code as a dictionary to convert the nucleotide sequence back to a protein sequence. Although there are, in principle, six different reading frames in which any DNA sequence can be read into protein (three on each strand), the correct one is usually recognized as the only one lacking frequent stop codons (see p. 108). The volume of DNA sequence information is already so large ($>10^6$ nucleotides) that computers must be used to store and analyze it.

Nucleic Acid Hybridization Reactions Provide a Sensitive Way of Detecting Specific Nucleotide Sequences[33]

When an aqueous solution of DNA is heated at 100°C or exposed to a very high pH (pH $\geq$ 13), the complementary base pairs that normally hold the two strands of the double helix together are disrupted and the double helix rapidly dissociates into two single strands. This process, called *DNA denaturation*, was for many years thought to be irreversible. However, in 1961 it was discovered that complementary single strands of DNA will readily re-form double helices (a process called **DNA renaturation** or **hybridization**) if they are kept for a prolonged period at 65°C. Similar hybridization reactions will occur be-

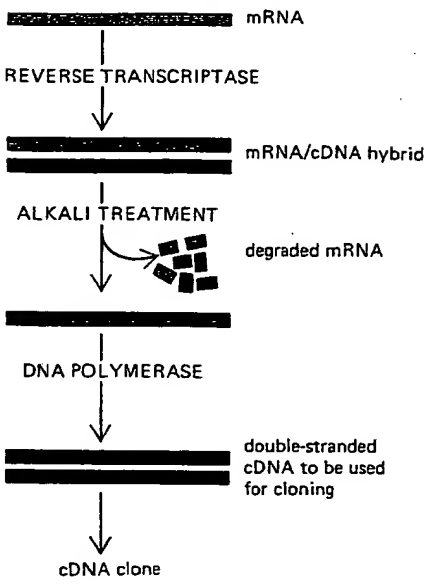

Figure 4–51 A DNA copy (cDNA) of an mRNA molecule is produced by the enzyme reverse transcriptase, a viral enzyme that uses an RNA strand as a template for the synthesis of a complementary DNA strand, thereby forming a DNA/RNA hybrid helix. Treatment of the DNA/RNA hybrid with alkali selectively degrades the RNA strand into nucleotides. The remaining single-stranded cDNA is then copied into double-stranded cDNA by the enzyme DNA polymerase.

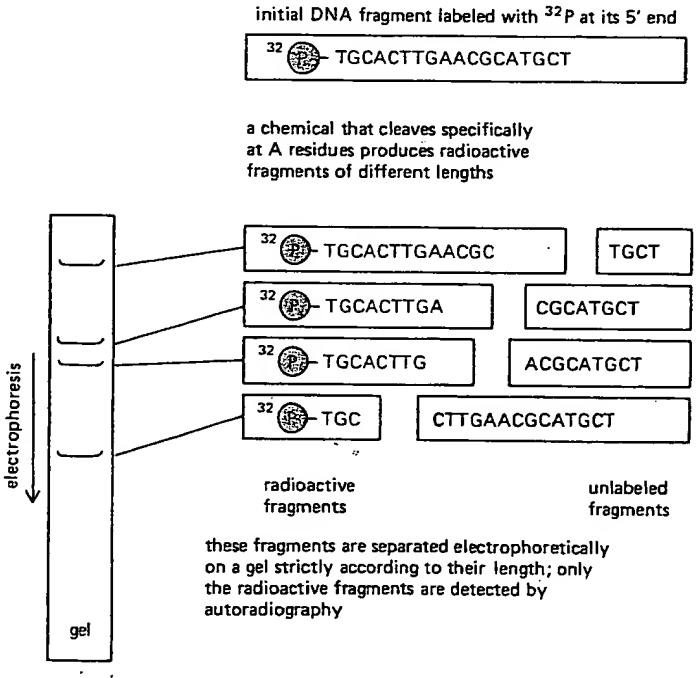

Figure 4–52 The generation of a family of DNA fragments by random cleavage of a DNA chain at a particular type of nucleotide. Each cleavage is produced by a mild chemical treatment that eliminates one nucleotide from the chain while leaving intact most of the nucleotides of the type eliminated. Only the left-hand fragments, possessing a 5′ terminal [^{32}P]phosphate group, are radioactive.

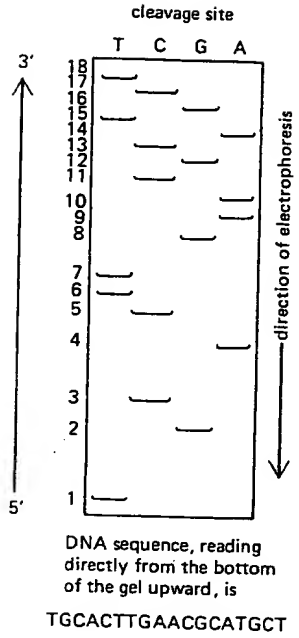

Figure 4–53 Schematic diagram showing one method for sequencing DNA. The type of procedure described in Figure 4–52 is carried out simultaneously on four separate samples of the same DNA using chemicals that cleave DNA specifically at T for the first sample, C for the second, G for the third, and A for the fourth. The resulting fragments are run in parallel lanes of the same gel, giving a pattern from which the DNA sequence is read. The nucleotide closest to the 5′ end of the sequence is determined by looking across the gel at level 1 (at the bottom of the gel) and seeing in which lane a band appears (T). The same procedure is repeated for level 2, then 3, and so on, to obtain the sequence. The method has been idealized here; in actuality the chemical treatments are less specific than shown.

tween any two single-stranded nucleic acid chains (DNA:DNA, RNA:RNA, or RNA:DNA), provided they have a complementary nucleotide sequence.

Because the rate of double-helix formation is limited by the rate at which two complementary nucleic acid chains happen to collide, the concentration of DNA molecules carrying a particular nucleotide sequence can be measured by the rate at which the DNA preparation of interest hybridizes to a radiolabeled cloned DNA probe of complementary sequence. This is such a stringent test that even complementary sequences present in a concentration of one molecule per cell can be detected (Figure 4–54). From such measurements it can be determined how many copies of the DNA sequence contained in the cloned probe are present in the DNA of a cell. While most sequences turn out to be present in only one or a few copies per haploid genome, others are present in hundreds of thousands of copies—the so-called *repeated DNA sequences.*

Alternatively, hybridization studies can be carried out with RNA isolated from cells to determine whether the DNA sequence that has been cloned is one of those transcribed into RNA and, if so, how many copies of the RNA are made per cell and in which types of cells and tissues. Somewhat more elaborate procedures identify the exact region of the cloned probe that hybridizes with cellular RNA molecules and thereby define the start and stop sites for RNA transcription (Figure 4–55); the regions that are cut out of the RNA transcripts during *RNA processing* (the intron sequences) are also identified in this way.

Radioactive cloned DNA probes are widely used to localize specific nucleic acid sequences in mixtures of DNA restriction fragments fractionated by gel electrophoresis. A replica of the gel is made by transferring all of the fractionated DNA fragments to a sheet of nitrocellulose paper either by diffusion or electrophoresis, a process called *blotting.* The locations of the fragments that hybridize to the radioactive DNA probe are then identified by autoradiography (Figure 4–56). In a similar way, nitrocellulose paper replicas can be made of crowded colonies of bacteria growing on an agar surface, so that hybridization of the paper with a specific radioactive probe can be used to identify the few cells carrying a newly cloned specific DNA fragment.

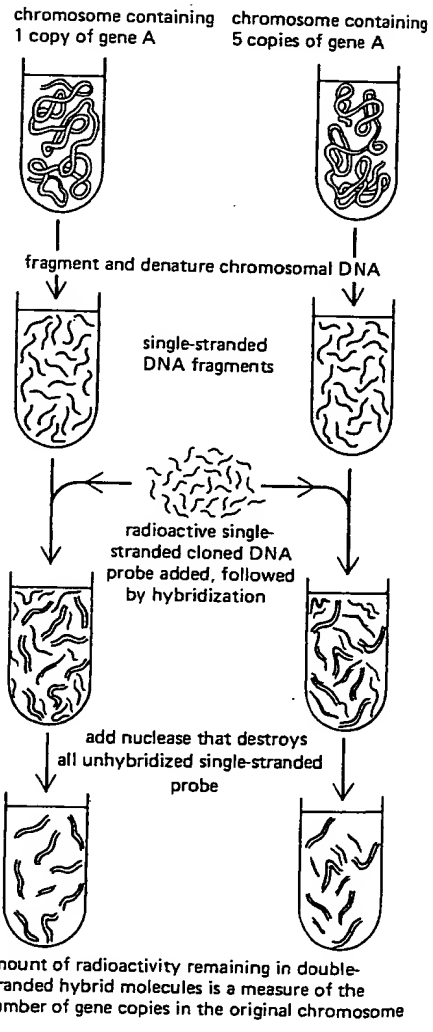

Figure 4–54 The measurement of the number of copies of a specific gene in a sample of DNA by means of DNA hybridization. The radioactive single-stranded DNA fragment used in such experiments is commonly referred to as a *DNA probe;* the chromosomal DNA is not radioactively labeled here.

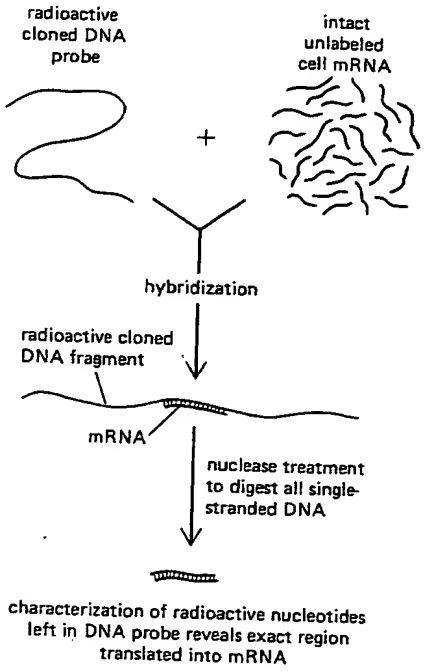

Figure 4–55 The use of nucleic acid hybridization to determine the region of a cloned DNA fragment that is transcribed into mRNA. The existence of intervening sequences (introns) in eucaryotic genes was discovered by this type of procedure.

In Situ Hybridization Techniques Are Used to Localize Specific Nucleic Acid Sequences in Chromosomes and Cells[34]

Nucleic acids, no less than other macromolecules, occupy precise positions within cells and tissues, and a great deal of potential information is lost when these molecules are extracted from cells by homogenization. For this reason, techniques have been developed in which nucleic acid probes are used in much the same way as labeled antibodies to localize specific nucleic acid sequences *in situ*, either in chromosomes or particular types of cells. In the original *in situ* hybridizations, highly radioactive nucleic acid probes were hybridized to squashed, fixed chromosomes that had been exposed briefly to a very high pH in order to disrupt their DNA base pairs. After extensive washing, the chromosomal regions that bound the radioactive probe were visualized by autoradiography (Figure 4–57). Recently, the spatial resolution of this technique has been improved by the development of special methods for labeling the nucleic acid probes with fluorescent dyes.

Similar *in situ* hybridization methods have been useful for detecting the presence of particular growing RNA transcripts on unusually large "lampbrush chromosomes"; here the chromosomes are not exposed to a high pH, so the chromosomal DNA itself remains double-stranded and thus cannot bind the probe. Comparable methods can be used on fixed tissue sections to determine which cells in a complex tissue contain cytoplasmic RNA molecules complementary to a particular DNA probe.

Recombinant DNA Techniques Allow Even the Minor Proteins of a Cell to Be Studied[31,35]

Until very recently, the only proteins that could be readily studied were relatively abundant components of the cell. Starting with several hundred grams of cells, a major protein—one that constitutes 1% or more of the total cellular protein—can be purified by a series of simple chromatographic and electro-

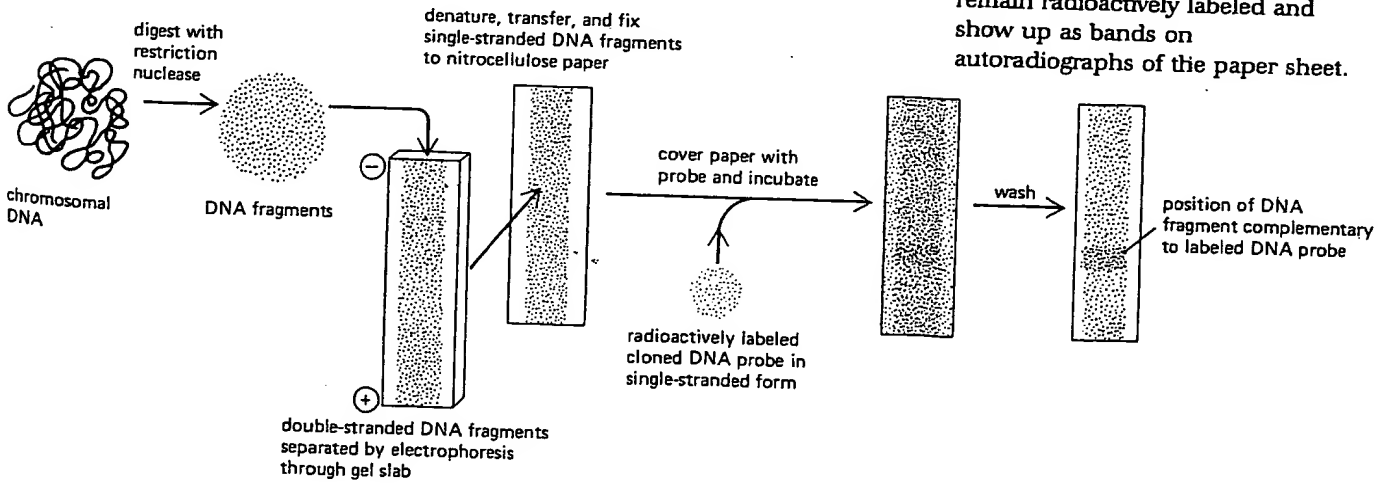

Figure 4–56 After a DNA sample has been cleaved by a restriction nuclease and then separated by electrophoresis, the many different DNA fragments present are transferred to nitrocellulose paper by blotting and then are exposed to a radioactive DNA probe for a prolonged period under annealing conditions. The sheet is washed extensively so that only those DNA fragments that hybridize to the probe remain radioactively labeled and show up as bands on autoradiographs of the paper sheet.

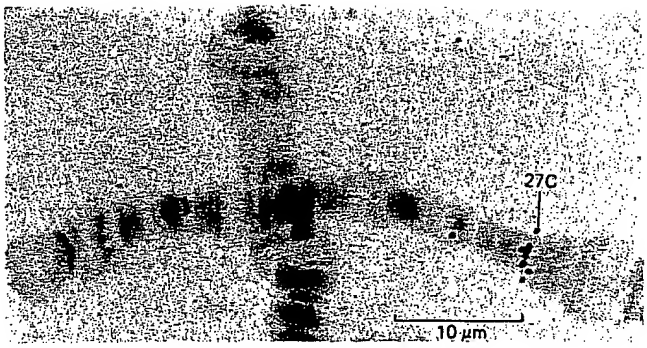

Figure 4–57 Localization of a *Drosophila* gene by *in situ* hybridization of a radioactive cloned DNA probe with *Drosophila* polytene chromosomes. The cluster of darkened silver grains detected in this autoradiograph is located at chromosome map position 27C, as indicated. Parts of two of the four giant chromosomes present in each larval salivary gland cell are shown. (Courtesy of Steven Henikoff.)

phoretic procedures to yield perhaps 0.1 g (100 mg) of pure protein. This quantity of protein is sufficient for conventional amino acid sequencing, detailed analysis of its biological or enzymatic activity (if any), and the production of antibodies, which can then be used to localize the protein in the cell. Moreover, if suitable crystals can be grown, the three-dimensional structure of the protein can be analyzed by x-ray diffraction crystallography. In this way, the structure and function of many abundant proteins have been determined, including hemoglobin, trypsin, immunoglobulin, and lysozyme.

The vast majority of the thousands of different proteins in a eucaryotic cell, including many of the most interesting ones, are present in only very small amounts. For most of them it is extremely difficult, if not impossible, to obtain more than a few micrograms of pure material. In principle, however, recombinant DNA technology has now made essentially any protein in the cell, including the minor ones, accessible to the same structural and functional studies that were previously possible only for a rare few. A summary of the steps that make this possible is given in Table 4–13.

Mutant Genes Can Now Be Made to Order[36]

Suppose one isolates a new protein from a cell extract and clones its gene by the "shotgun" technique described above. How can one discover what the protein does in the cell? The problem is surprisingly difficult since neither the three-dimensional structure of the protein nor the complete nucleotide sequence of its gene identifies the protein's function. And many proteins, such as structural components of the cell or proteins that are normally part of a large multienzyme complex, have no obvious activity when they are separated from the other components of the functional unit.

One approach already discussed is to inactivate the particular protein by means of a specific antibody. When combined with the technique of microinjection, this provides a powerful probe to test protein function. However, some antigenic sites on proteins will be inaccessible to antibody molecules even if the antibodies are injected into the cytoplasm. Furthermore, many antibodies bind to protein molecules without inactivating them.

Genetic approaches provide an elegant solution to this problem. Mutants that lack a particular protein or, more usefully, synthesize a temperature-

Table 4–13 Steps in the Purification of Large Amounts of a Minor Protein of the Cell Using Recombinant DNA Technology

1. Fractionate the cell extract by a series of conventional chromatographic procedures until the protein of interest is sufficiently enriched that a microgram can be obtained in pure form by cutting it out of a gel following high-resolution gel electrophoresis.
2. Analyze the denatured protein on a microsequenator to determine the sequence of the first 30 amino acids at its amino terminus.
3. Use the genetic code to predict the nucleotide sequences in mRNA corresponding to the above amino acid sequence. Using rapid chemical methods, synthesize a set of short DNA fragments, 15 to 20 nucleotides long, 1 of which will form complementary base pairs with part of the mRNA sequence. (There will be some ambiguity here since several different codons code for the same amino acid—p. 108).
4. Hybridize these short DNA fragments to total cellular mRNA and use them to direct reverse transcriptase to the mRNA molecules with complementary sequences. The reverse transcriptase then copies these complementary mRNA molecules to produce long cDNA molecules (Figure 4–51).
5. Produce large amounts of DNA containing the sequence of each of these cDNA molecules by cloning (Figures 4–49 and 4–50).
6. Hybridize DNA prepared from each cDNA clone to total cellular mRNA and thereby select and purify mRNA molecules that are complementary in sequence to each cloned cDNA sequence.
7. Translate each mRNA preparation obtained into protein by cell-free protein synthesis in order to determine which one codes for the desired protein.
8. Sequence the appropriate cDNA (Figure 4–53) and use the genetic code to determine the protein's complete amino acid sequence and where the coding sequence for the protein begins and ends.
9. Insert the cloned cDNA sequence into a specially engineered plasmid DNA vector containing inserted transcription and translation start signals. Use bacterial or yeast cells containing this new plasmid clone as the starting material for the isolation of large amounts (100 mg or more) of the purified protein.

sensitive version of the protein that is inactivated by a small increase (or decrease) in temperature may quickly reveal the function of the normal molecule. While this approach has been immensely useful, for example, in elucidating the principal metabolic pathways of bacteria, it has been mainly restricted to very rapidly replicating organisms—such as bacteria, yeast, nematode worms, and fruit flies—where very large numbers of mutants can be quickly isolated and then screened for a particular defect of interest.

In principle, the genetic approach can now be made more generally applicable by creating specific "mutations" outside the cell. With recently developed methods, a copy of an isolated cloned gene can be altered slightly by biochemical means and then put back into a cell, which now synthesizes an altered protein. In bacterial and yeast cells, this mutant gene will recombine with the normal gene often enough to make it possible to select for cells in which the mutant gene has replaced the single copy of the normal gene. In this way, cells carrying a specific protein in mutant form are made to order and the phenotype of a cell that lacks the normal gene thereby determined. Similar methods are not yet available for inserting a cloned mutant gene back into mammalian cells in place of the normal gene, but with the extraordinarily rapid rate of progress in recombinant DNA technology, it would not be surprising if this soon becomes possible.

Summary

Recombinant DNA technology has revolutionized the study of the cell. Any region of the cell's DNA can now be excised with restriction nucleases and produced in virtually unlimited quantities by DNA cloning and then sequenced at rates of hundreds of nucleotides a day. As a result, many genes and noncoding regions of the eucaryotic genome have already been sequenced.

By using nucleic acid hybridization methods, mRNA molecules corresponding to cloned DNA molecules can be detected, isolated, and translated into protein in cell-free systems. Furthermore, it is possible, in principle, to work backward from a protein to the gene that encodes it: by using a short stretch of amino acid sequence from the protein, specific DNA probes can be synthesized that will hybridize with the mRNA and DNA encoding the protein.

The practical consequences of recombinant DNA technology are far-reaching. Bacteria or yeast can be engineered to make a mammalian protein in virtually unlimited quantities, making it possible to analyze the structure and function of the protein or to use the protein as a vaccine or drug for medical purposes.

References

General

Cantor C.R.; Schimmel P.R. Biophysical Chemistry (3 vols.). San Francisco: Freeman, 1980. (A comprehensive account of the physical principles underlying many biochemical and biophysical techniques.)

Freifelder, D. Physical Biochemistry. San Francisco: Freeman, 1976.

Prescott, D., ed. Methods in Cell Biology. New York: Academic Press. (A multivolume series containing reviews of current techniques.)

Work, T.S.; Work, E.; Burden, R.H., eds. Laboratory Techniques in Biochemistry and Molecular Biology. Amsterdam: Elsevier/North-Holland Biomedical Press. (A multivolume series of practical guides to specialized biochemical procedures. Recent volumes include Sequencing of Proteins and Peptides, 1981; Gel Filtration Chromatography, 1980; and An Introduction to Affinity Chromatography, 1979.)

Cited

1. Bradbury, S. The Evolution of the Microscope. Elmsford, N.Y.: Pergamon, 1967.
   Bloom, W.; Fawcett, D.W. A Textbook of Histology, 10th ed. Philadelphia: Saunders, 1975. (A beautifully illustrated description of the anatomy of cells, as seen by light microscopy and transmission electron microscopy. Chapter 1 introduces the principal methods employed.)
2. Spencer, M. Fundamentals of Light Microscopy. Cambridge, Eng.: Cambridge University Press, 1982.
3. Nairn, R.C. Fluorescent Protein Tracing, 4th ed. New York: Churchill Livingstone, 1976.
   Lillie, R.D. Biological Stains, 8th ed. Baltimore: Williams & Wilkins, 1969.
4. Wischnitzer, S. Introduction to Electron Microscopy, 3rd ed. Elmsford, N.Y.: Pergamon, 1981.
   Weakley, B.S. A Beginner's Handbook in Biological Transmission Electron Microscopy, 2nd ed. New York: Churchill Livingstone, 1981.
5. Pease, D.C.; Porter, K.R. Electron microscopy and ultramicrotomy. *J. Cell Biol.* 91:287s–292s, 1981. (A short historical account.)
6. Everhart, T.E.; Hayes, T.L. The scanning electron microscope. *Sci. Am.* 226(1):54–69, 1972.
   Hayat, M.A. Introduction to Biological Scanning Electron Microscopy. Baltimore: University Park Press, 1978.
   Kessel, R.G. Tissues and Organs. San Francisco: Freeman, 1979. (An atlas of ver-

tebrate tissues seen by scanning electron microscopy.)

7. Pinto da Silva, P.; Branton, D. Membrane splitting in freeze-etching. *J. Cell Biol.* 45:598–605, 1970.
   Heuser, J. Quick-freeze, deep-etch preparation of samples for 3-D electron microscopy. *Trends Biochem. Sci.* 6:64–68, 1981.
8. Unwin, P.N.T.; Henderson, R. Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* 94:425–440, 1975.
9. Glusker, J.P.; Trueblood, K.N. Crystal Structure Analysis: A Primer. Oxford, Eng.: Oxford University Press, 1972.
   Kendrew, J.C. The three-dimensional structure of a protein molecule. *Sci. Am.* 205(6):96–111, 1961.
   Perutz, M.F. The hemoglobin molecule. *Sci. Am.* 211(5):64–76, 1964.
10. Paul, J. Cell and Tissue Culture, 5th ed. New York: Churchill Livingstone, 1975.
11. Harrison, R.G. The outgrowth of the nerve fiber as a mode of protoplasmic movement. *J. Exp. Zool.* 9:787–848, 1910. (Possibly the first use of tissue culture.)
12. Ham, R.G. Clonal growth of mammalian cells in a chemically defined, synthetic medium. *Proc. Natl. Acad. Sci. USA* 53:288–293, 1965.
    Hayashi, I.; Larner, J.; Sato, G. Hormonal growth control of cells in culture. *In Vitro* 14:23–30, 1978.
13. Harris, H.; Watkins, J.F. Hybrid cells derived from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature* 205:640–646, 1965.
    Ruddle, F.H.; Creagan, R.P. Parasexual approaches to the genetics of man. *Annu. Rev. Genet.* 9:407–486, 1975.
14. Colowick, S.P.; Kaplin, N.O., eds. Methods in Enzymology, Vols. 1–84. New York: Academic Press, 1955–1982. (A multivolume series containing general and specific articles on most procedures commonly employed in the biochemical analysis of cells.)
    Cooper, T.G. The Tools of Biochemistry. New York: Wiley, 1977.
    de Duve, C.; Beaufay, H. A short history of tissue fractionation. *J. Cell Biol.* 91:293s–299s, 1981.
15. Jovin, T.M.; Arndt-Jovin, D.J. Cell separation. *Trends Biochem. Sci.* 5:214–219, 1980.
    Herzenberg, L.A.; Sweet, R.G.; Herzenberg, L.A. Fluorescence-activated cell sorting. *Sci. Am.* 234(3):108–117, 1976.
16. de Duve, C. Exploring cells with a centrifuge. *Science* 189:186–194, 1975.
    Palade, G. Intracellular aspects of the process of protein synthesis. *Science* 189:347–358, 1975.
    Claude, A. The coming of age of the cell. *Science* 189:433–435, 1975.
    (A. Claude, C. de Duve, and G. Palade shared a Nobel Prize in 1974 for their work on tissue fractionation.)
    Meselson, M.; Stahl, F.W. The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 44:671–682, 1958. (Density gradient centrifugation was used to show the semiconservative replication of DNA.)
    Scheeler, P. Centrifugation in Biology and Medical Science. New York: Wiley, 1981.
17. Nirenberg, N.W.; Matthaei, J.H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* 47:1588–1602, 1961.
    Zamecnik, P.C. An historical account of protein synthesis, with current overtones—a personalized view. *Cold Spring Harbor Symp. Quant. Biol.* 34:1–16, 1969.
    Racker, E. A New Look at Mechanisms in Bioenergetics. New York: Academic Press, 1976. (Cell-free systems in the working out of energy metabolism.)
18. Andrews, A.T. Electrophoresis. New York: Oxford University Press, 1981. (A comprehensive guide to the theory, techniques, and biochemical applications of electrophoresis.)
19. O'Farrell, P.H. High resolution, two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007–4021, 1975.
20. Ingram, V.M. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* 178:792–794, 1956. (The original description of protein fingerprinting.)
    Cleveland, D.W.; Fischer, S.G.; Kirschner, M.W.; Laemmli, U.K. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* 252:1102–1106, 1977.

21. Walsh, K.A.; Ericsson, L.H.; Parmelee, D.C.; Titani, K. Advances in protein sequencing. *Annu. Rev. Biochem.* 50:261–284, 1981.
22. Rogers, A.W. Techniques of Autoradiography, 3rd ed. New York: Elsevier/North-Holland, 1979.
23. Calvin, M. The path of carbon in photosynthesis. *Science* 135:879–889, 1962. (A pioneer's account of one of the earliest uses of radioisotopes in biology.)
24. Coons, A.H. Histochemistry with labeled antibody. *Int. Rev. Cytol.* 5:1–23, 1956.
Hudson, L.; Hay, F.C. Practical Immunology, 2nd ed. Oxford, Eng.: Blackwell, 1980.
Eisen, H.N. Immunology, 3rd ed. New York: Harper & Row, 1981.
Anderton, B.H.; Thorpe, R.C. New methods of analyzing for antigens and glycoproteins in complex mixtures. *Immunol. Today* 2:122–127, 1980.
25. Milstein, C. Monoclonal antibodies. *Sci. Am.* 243(4):66–74, 1980.
Yelton, D.E.; Scharff, M.D. Monoclonal antibodies: a powerful new tool in biology and medicine. *Annu. Rev. Biochem.* 50:657–680, 1981.
26. Mueller, C.; Graessmann, A.; Graessmann, M. Microinjection: turning living cells into test tubes. *Trends Biochem. Sci.* 5:60–62, 1980.
Furusawa, M. Cellular microinjection by cell fusion: technique and applications in biology and medicine. *Int. Rev. Cytol.* 62:29–67, 1980.
27. Glover, D.M. Genetic Engineering: Cloning DNA. New York: Chapman and Hall, 1980. (A brief, 80-page summary of methodology.)
Williamson, R., ed. Genetic Engineering, Vols. 1–3. New York: Academic Press, 1979, 1981, 1982.
Watson, J.D.; Tooze, J. The DNA Story: A Documentary History of Gene Cloning. San Francisco: Freeman, 1981.
28. Smith, H.O. Nucleotide sequence specificity of restriction endonucleases. *Science* 205:455–462, 1979.
29. Novick, R.P. Plasmids. *Sci. Am.* 243(6):102–107, 1980.
Cohen, S.N. The manipulation of genes. *Sci. Am.* 233(1):24–33, 1975.
Maniatis, T.; et al. The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 15:687–701, 1978.
30. Maniatis, T.; Kee, S.G.; Efstratiadis, A.; Kafatos, F.C. Amplification and characterization of a β-globin gene synthesized *in vitro*. *Cell* 8:163–182, 1976.
31. Abelson, J.; Butz, E., eds. Recombinant DNA. *Science* 209:1317–1438, 1980. (A collection of articles by leaders in the field.)
32. Sanger, F. Determination of nucleotide sequences in DNA. *Science* 214:1205–1210, 1981.
Gilbert, W. DNA sequencing and gene structure. *Science* 214:1305–1312, 1981.
33. Hood, L.E.; Wilson, J.H.; Wood, W.B. Molecular Biology of Eucaryotic Cells: A Problems Approach, pp. 56–61, 192–201. Menlo Park, Ca.: Benjamin-Cummings, 1975. (Hybridization analyses clearly explained.)
Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517, 1975.
34. Pardue, M.L.; Gall, J.G. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. USA* 64:600–604, 1969.
Hennig, W. *In situ* hybridization of nucleic acids. *Trends Biochem. Sci.* 1:285–287, 1976.
35. Gilbert, W.; Villa-Komaroff, L. Useful proteins from recombinant bacteria. *Sci. Am.* 242(4):74–94, 1980.
Itakura, K. Synthesis of genes. *Trends Biochem. Sci.* 5:114–116, 1980.
36. Shortle, D.; Nathans, D. Local mutagenesis: a method for generating viral mutants with base substitutions in preselected regions of the viral genome. *Proc. Natl. Acad. Sci. USA* 75:2170–2174, 1978.
Hinnen, A.; Hicks, J.B.; Fink, G.R. Transformation of yeast. *Proc. Natl. Acad. Sci. USA* 75:1929–1933, 1978.
Berg, P. Dissections and reconstructions of genes and chromosomes. *Science* 213:296–303, 1981.
Anderson, W.F.; Diacumakos, E.G. Genetic engineering in mammalian cells. *Sci. Am.* 245(1):106–121, 1981.

Evidence Appendix G

0022-1767/93/1507-2992$02.00/0
The Journal of Immunology

Express Mail
Mailing Label No. EV832482014US

The Serine Protease Cofactor Factor V Is Synthesized by Lymphocytes[1]

Nancy L. L. Shen,[2]* Sao-Tah Fan,* Jayashree Pyati,† Richard Graff,† Robert J. LaPolla,† and Thomas S. Edgington[3]*

*Department of Immunology, The Scripps Research Institute, and †The R. W. Johnson Pharmaceutical Research Institute, La Jolla, CA 92037

ABSTRACT. Ag-specific cellular immune responses result in CD4+ T cell activation, which can induce the expression of tissue factor in cells of monocyte/macrophage lineage. This results in initiation of the coagulation protease cascade, with ultimate generation of thrombin. The latter is a potent and pleiotropic mediator of cellular responses and deposition of fibrin. To explore the requirements for extravascular cellular mediation of immune effector pathways, we have searched for a cellular source of the cofactor factor Va. Factor V mRNA was identified in human lymphoid cells by using reverse transcription followed by the polymerase chain reaction (RT-PCR). We confirmed our reverse transcription-polymerase chain reaction results by an independent cloning of factor V cDNA from a T cell cDNA library. The sequence of the factor V cDNA was virtually identical to hepatic factor V mRNA sequence. A limited span of mRNA, encoding part of the connecting region of the factor V protein, was found to contain nucleotide polymorphisms based on six nucleotide substitutions. Northern blot analysis confirmed the presence of a ~7-kb factor V mRNA in the Hut-78* human T lymphoma cell line and, at five- to eightfold less abundance, in unstimulated lymphocytes and long term allogeneic stimulated T cells. Immunocytology with factor V mAb identified factor V intracellularly in freshly isolated T lymphocytes but not on the surface of cells. These data provide evidence for factor V transcription and biosynthesis by human lymphocytes. They provide an additional perspective on how lymphocytes may contribute to inflammatory effector functions of cellular immune responses in extravascular sites through provision of cofactors necessary for the coagulation serine protease cascade. *Journal of Immunology*, 1993, 150: 2992.

Immune effector cell localization and local fibrin deposition are characteristic features of certain types of Ag-driven cellular immune tissue responses reflecting activation of the coagulation protease cascade (1, 2), notably in the classic delayed-type hypersensitivity lesions (3–6). Cells of monocyte lineage express TF[4] through cooperative interactions with Ag-stimulated CD4+ Th cells to initiate protease cascades (7, 8). TF, a cell surface receptor and catalytic cofactor, binds the serine protease factor VIIa, or its precursor zymogen factor VII, either from plasma (9–11) or via biosynthesis by macrophages (12) to form a functional cell surface TF·VIIa complex. This mediates limited proteolytic activation of factor X to the serine pro-

Received for publication April 14, 1992. Accepted for publication December 10, 1992.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] These studies were supported in part by Grants PO1 HL-16411 and P50 AI-27680 from the National Institutes of Health and were aided by a National Institutes of Health General Clinical Research Center grant. N. L. L. S. is supported by Training Grant HL 07195 from the National Institutes of Health. This is Publication 6902-IMM from the Department of Immunology of The Scripps Research Institute (La Jolla, CA).

[2] Current address: Gen-Probe Inc., 9880 Campus Point Drive, San Diego, CA 92122.

[3] Address correspondence and reprint requests to Thomas S. Edgington, M.D., The Scripps Research Institute (IMM-17), 10666 North Torrey Pines Road, La Jolla, CA 92037.

[4] Abbreviations used in this paper: Va, activated form of factor V; TF, tissue factor; HBS, HEPES buffered saline; RT, reverse transcriptase or transcription; PCR, polymerase chain reaction.

tease factor Xa (9–11). Factor Xa characteristically assembles with the cofactor factor Va^4 [fnc^4] for activation of prothrombin to thrombin (13, 14). Tracy et al. (15) observed that mononuclear leukocytes provide a suitable cell surface for the binding of factor Va, factor Xa, and calcium ions to assemble a functional binary complex, Va·Xa, the prothrombinase complex. The product thrombin is a recognized inflammatory mediator eliciting chemotaxis and cellular activation (16, 17) via the thrombin receptor (18), resulting in ion flux and second messenger generation. Thrombin also produces the local fibrin deposition (3–6, 19) characteristically observed in these cellular immune lesions and is associated with a concomitant increase in vascular permeability (19, 20).

Factor V is present in plasma at ~7 μg/ml and is also stored in α-granules of platelets (21). It is a large asymmetric and proteolytically labile glycoprotein (M_r 330,000) that is unlikely to be available and functional in the extracellular fluids. Factor V has latent cofactor function for factor Xa. Upon specific and limited proteolytic activation by thrombin or factor Xa, factor V is converted to factor Va (22, 23). Factor Va, a binding protein and cofactor for factor Xa that associates with cell surfaces, accelerates the rate of thrombin generation from prothrombin by more than 4 orders of magnitude, relative to the free factor Xa alone (13, 14). Factor V cDNA has been cloned from HepG2 cells (24) and human fetal liver (25, 26), and the complete factor V amino acid sequence and genomic organization have been elucidated (27).

There has been a description of factor V-immunoreactive molecules associated with human PBMC populations, appearing to be intracellular (28). A molecule functionally homologous to factor Va, designated as EPR-1, also has been described on the surface of myeloid and lymphoid cells (29, 30). To explore a source of legitimate factor Va cofactor function associated with immune effector cells, we have searched for factor V or similar molecules in human lymphocytes.

Materials and Methods

General methods

All chemicals were of reagent grade. All molecular biology techniques were performed by standard methods (31).

Cells

Peripheral venous blood from normal consenting donors (The Scripps Research Institute, National Institutes of Health General Clinical Research Center, with approval of the Institutional Human Research Committee) was collected in the presence of the anticoagulant heparin or acid citrate dextrose. The platelet-rich plasma was removed from whole blood after centrifugation at room temperature for 15 min at 180 × *g*. The buffy coat was collected and diluted with 2 volumes of RPMI 1640 medium (Irvine Scientific, Santa Ana, CA) and layered over Ficoll-Hypaque (Pharmacia, Piscataway, NJ) to isolate PBMC. Lymphocytes were isolated by centrifugation of PBMC on Sepracell-MN (Sepratech Corp., Oklahoma City, OK). T cells were isolated by nylon wool filtration of isolated lymphocytes with a 60 to 65% yield (32). Platelet contamination was <0.1 platelet/nucleated cell. Allogeneically stimulated T cells were prepared as described previously (30), except that the mixed lymphocyte cultures were stimulated unidirectionally every 7 days by irradiated allogeneic PBMC. Paired frozen human liver tissues and PBMC were obtained from patients with liver cirrhosis that had been referred to the University of Toronto, Toronto, Canada (courtesy of Dr. Gary A. Levy). Liver sections with relatively normal histology were used for isolation of RNA. Hut-78* (a subclone of the T lymphoma cell line Hut-78, ATCC TIB161), MLT, a human leukemia/lymphoma T cell line (a gift from Dr. Deno Dialynas, The Scripps Research Institute), and COS-1, a transformed African green monkey kidney cell line (ATCC CRL1650), were maintained in continuous culture in RPMI 1640 supplemented with 10% FCS (Gemini Bioproducts Inc., Calabasas, CA). HepG2, a human hepatocellular carcinoma cell line (ATCC HB8065), was grown in DMEM supplemented with 1% nonessential amino acids, 1% sodium pyruvate, and 10% FCS.

RNA isolation and Northern blot hybridization

Total cellular RNA was harvested from cell preparations by a guanidinium isothiocyanate-cesium chloride method (31). Poly(A)$^+$ was prepared by passage through an oligo-(dT) column (mRNA isolation kit; Pharmacia). For Northern blot analysis, 20 μg of total RNA were electrophoresed in a 1.0% agarose gel containing 6.6% formaldehyde. After electrophoresis RNA was blotted to nylon membranes (GeneScreen; New England Nuclear, Boston, MA) by capillary transfer in 10× SSPE (1× SSPE is 0.18 mM NaCl, 6.7 mM NaH_2PO_4, 6.7 mM Na_2HPO_4, 1 mM EDTA, pH 8.0). Two hybridization probes were used, 1) a factor V cDNA fragment (nucleotides 5414 to 5904; see *oligonucleotides*) isolated from a plasmid carrying a RT-PCR product, F2/F3, generated from factor V primer pair F2 and F3 (see Fig. 2) and 2) a 350-bp fragment for human GPIIIa (β_3) integrin encoding Ser-97 to Pro-219 (a gift from Joseph Loftus, The Scripps Research Institute). Both probes were labeled with [α-^{32}P]dATP by random priming (Boehringer Mannheim, Indianapolis, IN). Hybridizations were performed at 42°C in 6× SSPE, 1% (w/v) SDS, 10% dextran sulfate (Pharmacia), denatured salmon sperm DNA (100 μg/ml), 50% deionized formamide. Filters were washed twice at 65°C in 2× SSPE/0.5% SDS for 30 min. The final wash was in 0.1× SSPE at room temperature for 1 h. A human β-actin 27-base oligonucleotide (Clontech, Palo

Alto, CA) was labeled with [γ-^{32}P]ATP and used for estimation of this transcript as a control for RNA content on the membrane. Autoradiography of the washed filters was conducted on XAR-5 film (Eastman Kodak, Rochester, NY), by using an intensifying screen, at −70°C. Three to 4 days of development were typical for the factor V probe and platelet GPIIIa probe, with 1 day for the human β-actin probe.

Oligonucleotides

Two sets of 24-base oligonucleotides for RT-PCR were synthesized (Research Genetics, Huntsville, AL) on the basis of the primary sequence of human factor V (25, 26). The sense strands of factor V oligonucleotide primers were designated the "a" series, and the antisense primers were the "b" series. The translational initiation codon ATG was designated as +1. The position of each oligonucleotide is as follows: F1b, 6714 to 6737; F2a, 5881 to 5904; F2b, 5881 to 5904; F3a, 5414 to 5437; F3b, 5441 to 5464; F4a, 4917 to 4940; F4b, 4917 to 4940; F7a, 2776 to 2799; F7b, 2776 to 2799; F8a, 2092 to 2115; F9a, 1573 to 1596; F9b, 1588 to 1611; F10a, 1338 to 1361; F10b, 1330 to 1353; F14a, −25 to −2.

RT-PCR

For RT-PCR, RNA was rendered free of DNA with 0.3 U of DNase I (RNase-free grade; Boehringer Mannheim) at 37°C for 10 min, followed by phenol/chloroform (1:1, v/v) extraction and ethanol precipitation. First-strand cDNA was synthesized with a factor V antisense primer (b series, 0.25 pmol) from total RNA (1 to 2 μg) by using M-MLV RT (BRL, Gaithersburg, MD) or a cDNA synthesis kit (Invitrogen, San Diego, CA). The reactions were stopped by addition of an equal volume of 0.3 N NaOH/0.03 M EDTA, pH 8.0, and heating to 65°C for 1 h, followed by ethanol precipitation. The first-strand factor V cDNA in a mixture of paired factor V oligonucleotide primers (12.5 pmol each) was subjected to PCR amplification (33) by using Taql (Perkin Elmer Cetus, Norwalk, CT) in a DNA thermocyclic reactor (EpiCom, San Diego, CA). PCR amplification cycles were 94°C for 1 min, 55°C to 63°C for 2 min, and 72°C for 3 min for 25 cycles. Before the first cycle, the reaction mixture was denatured at 94°C for 10 min. After the last cycle, incubation at 72°C was extended for an additional 10 min. An aliquot of the reaction product was electrophoresed through a 2.0% agarose gel, and the DNA bands were visualized by ethidium bromide staining. In some experiments, the RT-PCR reaction product was precipitated with ethanol and digested with *Eco*RI (BRL, Gaithersburg, MD) before agarose gel electrophoresis.

cDNA library and plaque screening

A Hut-78* cDNA library was constructed by priming Hut-78* poly(A)$^+$ RNA (5 μg) with oligo(dT) for the synthesis of cDNA. The cDNA fragments greater than 1 kb were size selected and ligated into a λ gt10 vector modified at the *Eco*RI site by using a Librarian X cloning kit (Invitrogen, San Diego, CA). Plaques (5×10^5) were screened for factor V cDNA clones by using the F2/F3 fragments (Fig. 2), which were radiolabeled by random priming (oligolabeling kit; Pharmacia). Hybridization was performed in the presence of 50% formamide, 10× SSPE, 2× Denhardt's, 0.4% SDS, 100 μg/ml denatured salmon sperm DNA, at 42°C. The final wash of the filters was at 60°C in 0.1× SSC, 0.1% sodium sarcosine. The putative positive λ clones were digested with *Not*I, and the cDNA inserts were isolated, cloned in Bluescript SK2 (Stratagene, San Diego, CA), and isolated for sequencing.

Nucleotide sequencing and analysis

Amplified factor V cDNA fragments produced by RT-PCR were isolated from a 2% agarose gel and purified by using glass beads (GeneClean; Bio 101, San Diego, CA). The cDNA was treated with T4 polynucleotide kinase, blunt-ended with T4 DNA polymerase (New England Biolabs, Boston, MA), and cloned into the *Sma*I site of pUC18 that had been dephosphorylated. Recombinant clones (three to eight) of each RT-PCR product were selected by small-scale plasmid preparation and restriction enzyme digestion (31). Both strands of the factor V insert in each recombinant plasmid were sequenced by the dideoxynucleotide termination method. Sequencing was conducted on an automated DNA sequencer (Applied Biosystems model 370A) by using fluorescently labeled universal sequencing primers. For the nucleotide polymorphism studies, the recombinant plasmid carrying the F7/F8 insert was first digested with restriction enzyme to determine the orientation of insertion at the *Sma*I site. Sequencing was performed on DNA from two or three plasmids with the same orientation, mixed in equal molar ratios. Consensus sequence was obtained from multiple sequencing. The alignments with the hepatic factor V cDNA sequences (25, 26) were performed by using the GenBank data base entry and Intelligenetics programs. The sense strand of HTFV1 and HTFV4 (Fig. 1) were sequenced manually by using Sequenase II (United States Biochemicals, Cleveland, OH).

Production of mAb to human factor V

BALB/c mice were immunized i.p. with 25 μg of human factor V/Va (kindly provided by Dr. Daryl Fair, University of Texas, Tyler, TX, and isolated according to the method of Dahlback (22)) in CFA (0.2 ml), followed by 25-μg boosts of factor V i.p. The spleen cells from mice with a

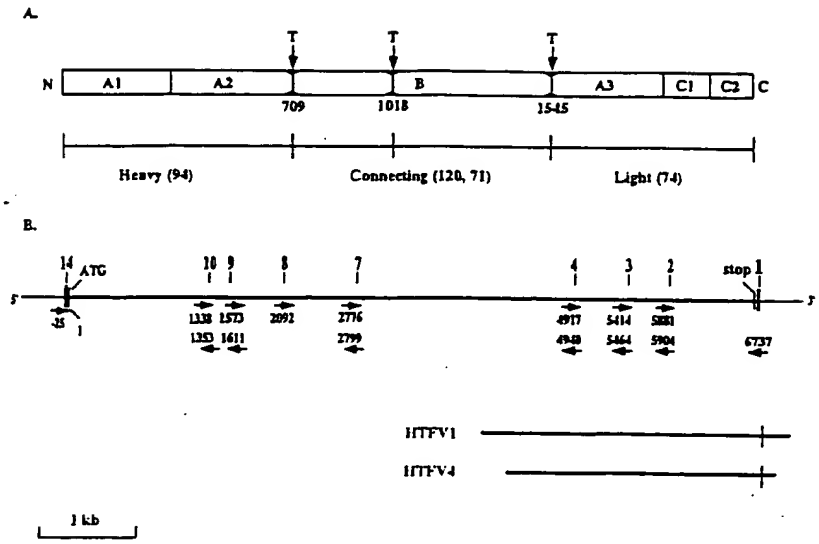

FIGURE 1. Factor V protein and cDNA. *A*, factor V protein (2196 amino acids) consists of three types of homologous domains (*A*, *B*, and *C*). Factor V is activated to factor Va by thrombin (*T*) or factor Xa through limited proteolysis, resulting in a H chain (amino acids 1 to 709, 94 kDa) and L chain (amino acids 1545 to 2196, 74 kDa), which associate by calcium-dependent hydrophobic interactions. *B*, oligonucleotide primers (1 through 14) for RT-PCR reactions are indicated *above the factor V message (7 kb) line. The numbers below the line* are the first nucleotides of each primer. The orientations of sense primers and antisense primers are indicated by the direction of the *arrows*. The translational start site ATG is designed as +1, and *stop* is the termination codon. Factor V cDNA clones (HTFV1 and HTFV4) are indicated.

serum titer of anti-factor V of greater than $1:10^5$ by dot blot were fused with P3×63Ag8.6 myeloma cells. The initial screening of clones producing antibody against factor V was done by solid phase RIA with ^{125}I-labeled V/Va and dot blot analysis with V/Va immobilized on nitrocellulose membranes. Clones producing mAb to factor V/Va were subcloned twice again by limiting dilution. All three selected clones produced IgG1 antibody by isotyping (Zymed Laboratories, San Francisco, CA). The hybridomas were grown in HT medium, 5×10^6 cells were inoculated i.p. into Pristane-primed BALB/c mice, and IgG1 mAb was purified from ascites fluid by using Affigel-protein A (Bio-Rad, Richmond, CA). The anti-factor V mAb 8C4 (IgG1), 15B6 (IgG1), and 22D6 (IgG1) were biotinylated (Calbiochem, San Diego, CA). The mAb that were selected do not neutralize factor Va coagulant activity of plasma and do not react with the monocytic cell line THP-1, in contrast to the EPR-1 mAb (29, 30). By Western blot all three mAb reacted with factor V, intermediate activation products, and purified factor Va L chain but not with H chain (kindly provided by Dr. Kenneth Mann, University of Vermont).

Immunocytology

Immunofluorescence reactions were performed on ice. Freshly isolated T cells were fixed for 5 min in HBS (0.9% sodium chloride, 20 mM HEPES, pH 7.4, 5 mM $CaCl_2$) containing 1% (w/v) paraformaldehyde. After one HBS wash, fixed cells were reacted for 30 min with the designated biotinylated mAb in HBS containing 2.5% BSA, (Sigma Chemical Co., St. Louis, MO) and 0.1% Triton X-100 (Sigma). Cells were washed once and exposed for 30 min to FITC-conjugated streptavidin (Becton Dickinson, San Jose, CA) diluted 1/25 with HBS containing 2.5% BSA. The biotin-conjugated mAb were used at 3 μg/ml final. mAb 22D6, 8C4, and 15B6 were used either mixed in equal amounts in a pool or individually. Anti-human TF mAb 1E7 (IgG1) was used as an isotype control. The irrelevant mAb HB3 (IgG2a) to anti-SV40 large T viral Ag served as an additional negative control. Reacted cells were spun onto glass slides (Cytospin; Shandon Southern Instruments, Camberly, Surrey, UK), mounted with Gelmount (Biomeda, Foster City, CA), and observed with a Nikon photomicroscope (Nikon Instruments, San Diego, CA) equipped for epifluorescence. In some experiments, intact peripherial blood-derived T cells were reacted with mAb 22D6, 8C4, and 15B6.

Results

RT-PCR cloning of factor V in Hut-78*

To analyze the presence of factor V transcripts in T cells, we first investigated the Hut-78* cell line. A PCR (33) modified for RNA analysis (RT-PCR) was used to identify putative factor V transcripts. A series of oligonucleotide primers based on the published hepatic factor V cDNA are illustrated in Figure 1. Each antisense b oligonucleotide primer and Hut-78* total RNA were used in the presence of RT to synthesize the first-strand cDNA. Paired primers,

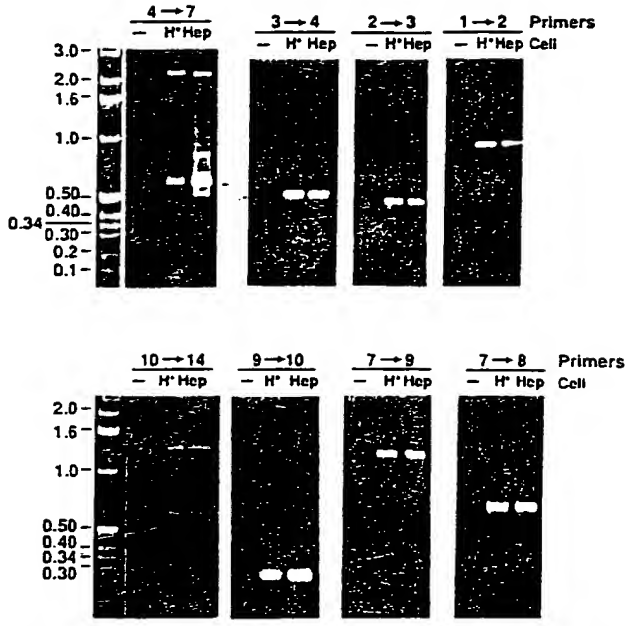

FIGURE 2. RT-PCR-amplified fragments of factor V cDNA. First-strand cDNA synthesis of Hut-78* RNA (1 μg) was either in the presence of RT (*H*⁺) or in the absence of RT (–). HepG2 RNA (1 μg) (*Hep*) was the positive control. The PCR that followed used the indicated pairs of factor V primers. The PCR products (10 μl of a 100-μl reaction) were visualized by ethidium bromide staining of agarose gels. Molecular size markers (1-kb ladder) are shown.

F1 and F2, F2 and F3, and F3 and F4, were used to specifically amplify factor V or related transcripts encoding the L chain region and primer pairs. F4 and F7, F7 and F8, F7 and F9, F9 and F10, and F10 and F14 were used for the amplification of the factor V message encoding the connecting and H chain regions. Exploring conditions of hybridization of varying stringency, we readily observed products of sizes appropriate for factor V that were indistinguishable from those produced by analysis of RNA from the positive control HepG2 cells (Fig. 2). In addition, a 0.6-kb DNA fragment that was shorter than the expected 2.2-kb F4/F7 fragment was observed with RNA from the Hut-78* and HepG2 cells. Sequencing of the 0.6-kb fragment showed sparse sequence similarity with factor V, presumably generated from nonspecific priming of an unrelated transcript in the RT-PCR reactions.

The appropriate size cDNA bands from Hut-78* cells were cloned into pUC18 and individual recombinant plasmids were selected. Multiple clones carrying recombinant plasmids from each segment of amplified factor V cDNA were isolated and the plasmid DNA inserts were sequenced individually. The sequences determined from three to 14 different recloned plasmids provided a consensus sequence for each product. No clones contained deletion or insertion mutations. They agreed with the factor V mRNA sequence from HepG2 and fetal liver cDNA (25, 26), with the exception of the factor V cDNA amplified by using primer pairs F7 and F8.

Both strands from 14 independent recombinant plasmid clones carrying the amplified F7/F8 factor V cDNA fragment were sequenced. Six nucleotide base substitutions were identified (Fig. 3). Two substitutions, from thymine to cytosine and from cytosine to thymine at positions 2209 and 2236, respectively, were silent mutations. The remaining four guanine to adenine base substitutions resulted in a silent mutation at position 2302 and amino acid changes from arginine to lysine at position 2573, from arginine to histidine at position 2595, and from glutamic acid to lysine at position 2773. These deduced amino acid changes are conservative substitutions and may not significantly affect factor V function or only subtly. In addition, half of the clones (seven of 14) have an adenine to guanine substitution at position 2290, a silent substitution, which abolished the *Eco*RI site.

Factor V cDNA library cloning

To confirm and extend the aforementioned observations, a Hut-78* λ gt10 cDNA library was independently constructed and probed with radiolabeled 0.5-kb factor V cDNA fragment F2/F3 generated by RT-PCR from Hut-78* cells. Screening of 5×10^5 recombinant phage gave six positive clones in the primary assay screen. Only the two longest were further characterized; one contained a 3.0-kb insert and the other contained a 2.7-kb insert. These inserts were subcloned in Bluescript to generate plasmids HTFV1 and HTFV4, respectively (Fig. 1). Sequencing of HTFV1 and HTFV4 revealed identity with the hepatic factor V cDNA sequence (25, 26). The translational stop codon and the poly(A) tail were conserved in the Hut-78* factor V transcript, and no putative transmembrane domain sequence was present.

Detection of factor V transcript in lymphocytes

To compare the relative abundance of factor V transcripts in Hut-78* cells with that in HepG2 cells, total RNA was probed with a radiolabeled factor V cDNA fragment F2/F3. As shown in Figure 4*A*, a 7-kb band was detected by Northern blot hybridization, in agreement with the size and abundance of the factor V mRNA from HepG2 cells. The abundance of mRNA was corrected relative to β-actin transcripts.

Further, we inquired whether the expression of factor V mRNA is normal for T cells, rather than an aberrant characteristic of the transformed T cell line Hut-78*. We performed Northern blot hybridization to analyze for factor V transcripts in lymphocytes isolated from healthy individuals. As illustrated in Figure 4*B*, a discrete factor V transcript was identified in RNA from both freshly isolated lymphocytes and long term allogeneically stimulated T

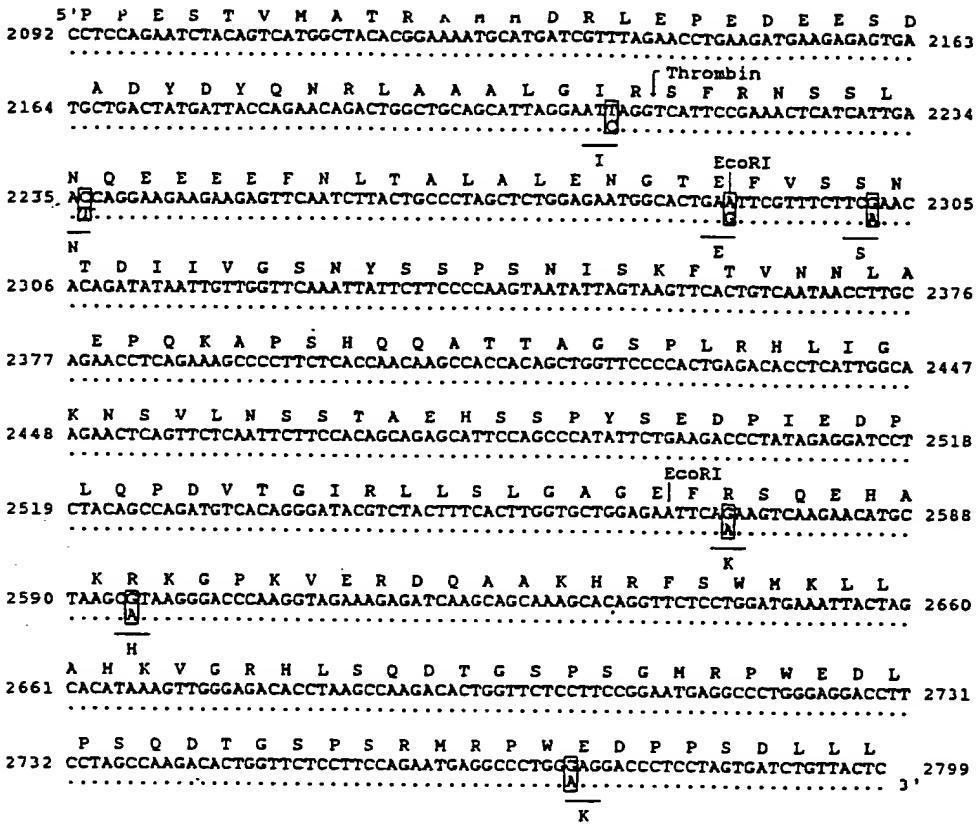

FIGURE 3. Factor V cDNA polymorphisms. Six novel nucleotide substitutions in the span of 2092 to 2799 and one nucleotide mutation at the *Eco*RI site are shown *below* and *boxed* with the published factor V cDNA sequence. The predicated amino acid changes are indicated in single-letter code below the nucleotide substitutions. Sequence differences are only indicated.

cells. It was ~7 kb in length and co-migrated with the major species found in Hut-78* and HepG2 cells. The hybridization signals of lymphocytes and allogeneically stimulated T cells were five- to eightfold less abundant, relative to Hut-78* transcripts, by densitometer scanning. Factor V mRNA could not be identified in the MLT cell line ($CD4^+CD8^+$) or an irrelevant monkey kidney cell line (COS-1), indicating that not all T cell lines or other cells express factor V mRNA. The integrity and amount of RNA samples before hybridization are shown in Figure 4*C*. To rule out platelet contamination as a source of contaminating factor V transcripts, four lines of evidence are available. First, the T cell lymphoma line Hut-78* was positive and cannot be contaminated by platelets. Second, T cells in prolonged in vitro culture with repeated long term allogenic stimulation were positive, which provides evidence for factor V transcripts in nontransformed activated lymphocytes. Third, the Northern blots of lymphocyte RNA (Fig. 4*B*) with radiolabeled probe for GPIIIa, one of the most abundant mRNA species in platelets, gave no signal (data not shown). GPIIIa transcripts have been demonstrated in platelets by RT-PCR (34) and Northern blots (35). Fourth, immunocytologic analysis (see below) demonstrated immunoreactive factor V in most lymphocytes.

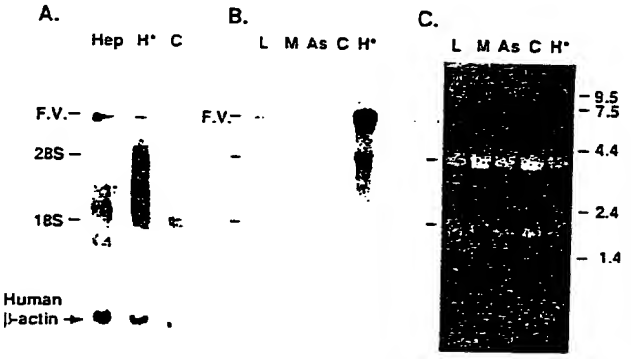

FIGURE 4. Factor V mRNA in T cells by Northern blot. *A*, total RNA (20 μg) from HepG2 (*Hep*), Hut-78* (*H**), and COS-1 (*C*) cells was transferred from a 1.0% formaldehyde-agarose gel to a nylon membrane (GeneScreen) and probed for the presence of factor V transcripts by using radiolabeled factor V cDNA fragment F2/F3. *F.V.*, factor V mRNA (7 kb). The filter was further probed with human β-actin as an internal control, shown at *bottom*. *B*, total RNA (20 μg) isolated from various lymphocyte preparations was probed with a radiolabeled factor V cDNA fragment described in *A*. *Lane L*, lymphocytes; *lane M*, MLT cells; *lane As*, allostimulated T cells; *lane C*, COS-1 cells; *lane H**, Hut-78* cells. *C*, RNA shown in *B* were visualized in a formaldehyde-agarose gel and stained with ethidium bromide before hybridization. The molecular size markers and 28S and 18S rRNA are indicated.

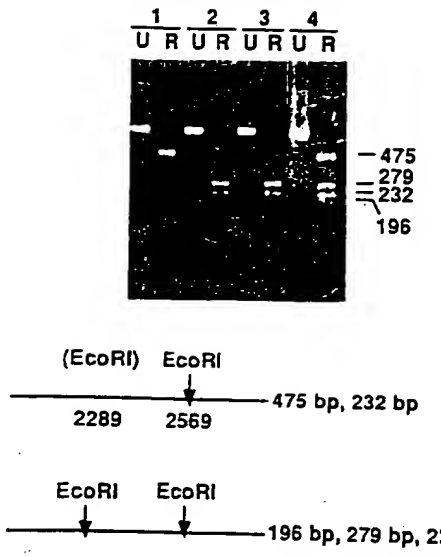

FIGURE 5. *Eco*RI RFLP. Factor V cDNA (F7/F8) fragments were prepared by RT-PCR from lymphocyte RNA (2 μg) of four individuals. It was precipitated with ethanol and divided; half was digested with *Eco*RI (*R*) and half was undigested (*U*). DNA fragments were separated on a 2% agarose gel and stained with ethidium bromide. The sizes of the restriction fragments generated by two *Eco*RI sites are 232 bp, 279 bp, and 196 bp. The *Eco*RI site mutation (in parentheses) at position 2289 resulted in 475-bp and 232-bp fragments.

Nucleotide sequence variation in factor V transcript

Lymphocyte RNA from six healthy individuals was assayed for the novel nucleotide substitutions that were identified in Hut-78*. The amplified RT-PCR products obtained by using primer pairs F7 and F8 were subjected to *Eco*RI enzyme digestion. Figure 5 illustrates three representative patterns of *Eco*RI site polymorphism at position 2289. Three bands, of 279 bp, 232 bp, and 196 bp in length, were identified in individuals having two *Eco*RI sites (Fig. 5, *lanes 2R* and *3R*). Two bands, of 475 bp and 232 bp in length, were found for individuals having only the *Eco*RI site at position 2569, with a loss of the *Eco*RI site at position 2289 (Fig. 5, *lane 1R*). Four bands, of 475 bp, 279 bp, 232 bp, and 196 bp in length, indicated individuals having two types of cDNA, one containing two *Eco*RI sites and the other with only one *Eco*RI site at position 2569 (Fig. 5, *lane 4R*), which may be derived from heterozygote alleles. Sequencing of cloned F7/F8 factor V cDNA from lymphocyte RNA indicated that the six novel nucleotide substitutions (Fig. 3) were observed in five of the six individuals studied, and loss of the *Eco*RI site at nucleotide 2290 is due to an adenine to guanine base substitution. To examine the possibility that these substitutions were lymphocyte-specific post-transcriptional nucleotide modifications, we analyzed paired samples of liver tissue and PBMC from the same individual. The same RT-PCR procedure was applied to two individual samples. Sequencing indicated that the patterns of the six nucleotide substitutions were precisely conserved in RNA from both liver cells and PBMC.

Cellular factor V

The translational product of factor V mRNA was studied in freshly isolated peripheral T lymphocytes. These cells were fixed, either permeabilized or not, and analyzed with specific anti-human factor V mAb (22D6, 8C4, and 15B6) reactive with factor V/Va L chain. Immunofluorescence microscopy illustrates (Fig. 6*C*) that immunoreactive factor V molecules are observed in T cells in condensed interconnecting filamentous structures, suggesting possible concentration in the Golgi, with occasional cells containing high levels of factor V (Fig. 6*D*). There was evident intracellular factor V/Va Ag in approximately 90% of T lymphocytes. Each of the three V/Va mAb gave identical results (data not shown). There were no detectable signals with irrelevant mAb HB3 (IgG2a) or IE7 (IgG1) (Fig. 6*B*). No factor V/Va Ag was detected on the surface of cells (Fig. 6*A*).

Discussion

Serine proteases, including the coagulation, complement, and granzymes are effectors of inflammatory responses locally invoked in Ag-specific immune responses. A central feature of the coagulation and complement cascades is the utilization of protein cofactors for cell surface localization, amplification of catalytic activity, and mediation of specific substrate recognition. The initiating receptor and cofactor TF is induced by Ag-driven Th cells (7, 8) and exhibits these characteristics. However, to propagate the coagulation protease response factor V/Va or a functionally homologous cofactor is required. We established the presence of legitimate transcripts of the precofactor of factor V in T cells. The factor V mRNA is correctly spliced from RT-PCR analysis and of the correct size by Northern blot analysis. There are no deletions, insertions, alternative splicing, or alteration of the translational reading frame determined by sequence analysis. Further, specific mAb localize factor V molecules within the T cells by immunofluorescence microscopy, indicating that the mRNA is translated to protein.

Our data indicate that the abundance of factor V transcripts in freshly isolated lymphocytes is five- to eight-fold less than that in Hut-78* T cells or HepG2 cells, consistent with relatively modest levels of expression (Fig. 4). This may explain the failure heretofore to detect factor V transcripts in total PBL by Northern blot analysis (25). Use of RT-PCR permitted direct restriction enzyme digestion and sequence analysis of the factor V transcript of lymphocytes. When various individuals were analyzed, a novel set of nucleotide polymorphisms were demonstrated at the region of the factor V transcript encoding the connecting

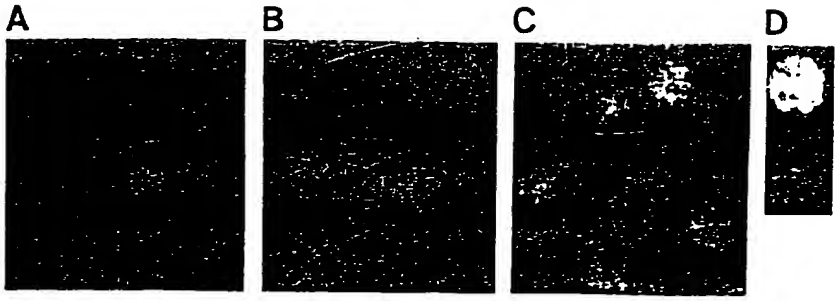

FIGURE 6. Immunocytologic analysis of T cells for factor V. Freshly isolated peripheral blood T cells were reacted with factor V mAb (8C4, 22D6, or 15D6). Either intact cells (*A*) or permeabilized cells (*C*) were used. No reaction was observed for irrelevant mAb (HB3 or IF7) on the permeabilized cells (*B*). *D*, strong reaction for factor V. All panels used 2000× magnification with identical exposure at the same light intensity.

region. Restriction digestion of the RT-PCR product establishes that individuals are either homozygous or heterozygous for an *Eco*RI restriction site. The origin of these nucleotide variations in factor V transcripts is most likely attributed to individual genetic polymorphism, because both lymphocyte and liver from the same individuals had the same patterns of nucleotide changes, excluding the possibility of tissue-specific post-transcriptional nucleotide modifications (36, 37).

Expression of factor V has been established for only limited cell types, including human hepatoma (38), bovine endothelial and smooth muscle cells (39, 40), and rabbit macrophages (41), where it is secreted. Additional studies have demonstrated factor V-immunoreactive molecules that are associated with cells including guinea pig megakaryocytes (42), lysed human PBMC (28), and human megakaryocytes (43). In the present study we identified T cell-associated factor V-immunoreactive molecules by using anti-human factor V mAb that exhibit no detectable cross-reactivity with EPR-1 (30). We found expression of factor V mRNA by Hut-78*, a T lymphoma cell line, which clearly excludes derivation from other cells or factor V endocytosed from plasma. The evolution of the currently known functionally related molecules, including EPR-1, the tumor-associated factor Xa receptor (44), and the recently cloned murine mammary epithelial cell surface molecule (45), remains to be elucidated.

Intracellular sequestration of factor V in T cells not only might provide protection from degradation of this large protease-sensitive protein but also might deliver it to extravascular sites of lymphocyte traffic, similar to delivery of granzymes to the surface of target cells by cytolytic T cells (46). We speculate that mediators involved in T cell activation and collaboration in cellular response may elicit the local release of factor V from lymphocytes. The activation of factor V by factor Xa and formation of the binary Va·Xa prothrombinase complex on local cell surfaces would profoundly enhance local generation of thrombin.

Serine proteases are recognized as potentially important for the effector phase of the immune response. The granzyme family of serine proteases are expressed in thymocytes and activated mature T cells and participate in lytic events by as yet obscure mechanisms (47, 48). Further, evidence exists that serine protease activity may be required for optimal lymphokine release by T cells (49). These observations, coupled with primary sequence homology between certain T cell serine proteases and coagulation serine proteases (50), lead to a unifying hypothesis. Because the coagulation protease cascade utilizes cofactors for cell surface localization, enhancement, and regulation of functional proteolytic activity, the possibility should be considered that the participation of cofactors for cell surface assembly of cellular serine proteases may represent a general paradigm.

Acknowledgments

We wish to thank Dr. Dario C. Altieri (Scripps Research Institute) for providing Hut-78* and allogeneic stimulated T cells and Dr. Deno Dialynas (Scripps Research Institute) for providing MLT cells. Dr. Gary A. Levy at the University of Toronto (Toronto, Canada) provided paired frozen liver tissue and PBMC. Dr. Kenneth Mann provided human factors V and Va and L and H chains. Dr. Ingrid Schraufstatter and Zenaida Oades shared human venous blood specimen samples obtained from General Clinical Research Center (Scripps Research Institute). The authors thank Drs. Xiaoping Du and Joseph Loftus for discussion, Bruce Fowler and Jennifer Royce for excellent technical assistance, and Barbara Parker for preparation of the final manuscript.

References

1. Davie, E. W., K. Fujikawa, and W. Kisiel. 1991. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry 30:10363.*
2. Edgington, T. S., N. Mackman, S.-T. Fan, and W. Ruf. 1992. Cellular immune and cytokine pathways resulting in tissue factor expression and relevance to septic shock. *Nouv. Rev. Fr. Hematol. 34(Suppl):S13.*
3. Colvin, R. B., R. A. Johnson, M. C. Mihm, Jr., and H. F. Dvorak. 1973. Role of the clotting system in cell-mediated hypersensitivity. I. Fibrin deposition in delayed skin reactions in man. *J. Exp. Med. 138:686.*
4. Colvin, R. B., and H. F. Dvorak. 1975. Role of the clotting system in cell-mediated hypersensitivity. II. Kinetics of fibrinogen/fibrin accumulation and vascular permeability

changes in tuberculin and cutaneous basophil hypersensitivity reactions. *J. Immunol. 114:377.*

5. Colvin, R. B., M. W. Mosesson, and H. F. Dvorak. 1979. Delayed-type hypersensitivity skin reactions in congenital afibrinogenemia lack fibrin deposition and induration. *J. Clin. Invest. 63:1302.*
6. Hopper, K. E., C. L. Geczy, and W. A. Davies. 1981. A mechanism of migration inhibition in delayed-type hypersensitivity reactions. I. Fibrin deposition on the surface of elicited peritoneal macrophages *in vivo*. *J. Immunol. 126:1052.*
7. Gregory, S. A., and T. S. Edgington. 1985. Tissue factor induction in human monocytes: two distinct mechanisms displayed by different alloantigen-responsive T cell clones. *J. Clin. Invest. 76:2440.*
8. Fan, S. T., A. L. Glasebrook, and T. S. Edgington. 1990. Clonal analysis of CD4 $^+$ T helper cell subsets that induce the monocyte procoagulant response. *Cell. Immunol. 128:52.*
9. Broze, G. J. 1982. Binding of human factor VII and VIIa to monocytes. *J. Clin. Invest. 70:526.*
10. Bach, R., R. Gentry, and Y. Nemerson. 1986. Factor VII binding to tissue factor in reconstituted phospholipid vesicles: induction of cooperativity by phosphatidylserine. *Biochemistry 25:4007.*
11. Ploplis, V. A., T. S. Edgington, and D. S. Fair. 1987. Initiation of the extrinsic pathway of coagulation: association of factor VIIa with a cell line expressing tissue factor. *J. Biol. Chem. 262:9503.*
12. Tsao, B. P., D. S. Fair, L. K. Curtiss, and T. S. Edgington. 1984. Monocytes can be induced by LPS triggered T lymphocytes to express functional factor VII/VIIa protease activity. *J. Exp. Med. 159:1042.*
13. Miletich, J. P., C. M. Jackson, and R. W. Majerus. 1978. Properties of the factor Xa binding site on human platelets. *J. Biol. Chem. 253:6908.*
14. Nesheim, M. E., J. B. Taswell, and K. G. Mann. 1979. The contribution of bovine factor V and factor Va to the activity of prothrombinase. *J. Biol. Chem. 254:10952.*
15. Tracy, P. B., L. L. Eide, and K. G. Mann. 1985. Human prothrombinase complex assembly and function on isolated peripheral blood cell populations. *J. Biol. Chem. 260:2119.*
16. Bar-Shavit, R., A. Kahn, and G. D. Wilner. 1983. Monocyte chemotaxis: stimulation by specific exosite region in thrombin. *Science 220:728.*
17. Hattori, R., K. K. Hamilton, R. D. Fugate, R. P. McEver, and P. J. Sims. 1989. Stimulated secretion of endothelial von Willebrand factor is accompanied by rapid redistribution to the cell surface of the intracellular granule membrane protein GMP-140. *J. Biol. Chem. 264:7768.*
18. Vu, T.-K. H., D. T. Hung, V. I. Wheaton, and S. R. Coughlin. 1991. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell 64:1057.*
19. Dvorak, H. F., S. J. Galli, and A. M. Dvorak. 1986. Cellular and vascular manifestations of cell-mediated immunity. *Hum. Pathol. 17:122.*
20. Clauss, M., M. Gerlach, H. Gerlach, J. Brett, F. Wang, P. C. Familletti, Y.-C. E. Pan, J. V. Olander, D. T. Connolly, and D. Stern. 1990. Vascular permeability factor: a tumor-derived polypeptide that induces endothelial cell and monocyte procoagulant activity, and promotes monocyte migration. *J. Exp. Med. 172:1535.*
21. Tracy, P. B., X. Eidell, E. J. W. Bowie, and K. G. Mann. 1982. Radioimmunoassay of factor V in human plasma and platelets. *Blood 60:59.*
22. Dahlback, B. 1980. Human coagulation factor V purification and thrombin-catalyzed activation. *J. Clin. Invest. 66:583.*
23. Esmon, C. T. 1979. The subunit structure of thrombin-activated factor V, isolation of activated factor V, separation of subunits, and reconstitution of biological activity. *J. Biol. Chem. 254:964.*
24. Kane, W. H., and E. W. Davie. 1986. Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. *Proc. Natl. Acad. Sci. USA 83:6800.*
25. Jenny, J. R., D. D. Pittman, J. J. Toole, R. W. Kriz, R. A. Aldape, R. M. Hewick, R. J. Kaufman, and K. G. Mann. 1987. Complete cDNA and derived amino acid sequence of human factor V. *Proc. Natl. Acad. Sci. USA 84:4846.*
26. Kane, W. H., A. Ichinose, E. S. Hagen, and E. W. Davie. 1987. Cloning of cDNAs coding for the heavy chain region and connecting region of human factor V, a blood coagulation factor with four types of internal repeats. *Biochemistry 26:6508.*
27. Cripe, L. D., K. D. Moore, and W. H. Kane. 1992. Structure of the gene for human coagulation factor V. *Biochemistry 31:3777.*
28. Tracy, P. B., M. S. Rohrbach, and K. G. Mann. 1983. Functional prothrombinase complex assembly on isolated monocytes and lymphocytes. *J. Biol. Chem. 258:7264.*
29. Altieri, D. C., and T. S. Edgington. 1989. Sequential receptor cascade for coagulation proteins on monocytes: constitutive biosynthesis and functional prothrombinase activity of a membrane form of factor V/Va. *J. Biol. Chem. 264:2969.*
30. Altieri, D. C., and T. S. Edgington. 1990. Identification of effector cell protease receptor-1, a leukocyte-distributed receptor for the serine protease factor Xa. *J. Immunol. 145:246.*
31. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
32. Falkoff, R. M., M. Peters, and A. S. Fauci. 1982. T cell enrichment and depletion of human peripheral blood mononuclear cell preparations: unexpected findings in the study of the functional activities of the separated populations. *J. Immunol. Methods 50:39.*
33. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science 239:487.*
34. Newman, P. J., J. Gorski, G. C. White II, S. Gidwitz, C. J. Cretney, and R. H. Aster. 1988. Enzymatic amplification of platelet-specific messenger RNA using the polymerase chain reaction. *J. Clin. Invest. 82:739.*
35. Bray, P. F., and M. A. Shuman. 1990. Identification of an abnormal gene for the GPIIIa subunit of the platelet fibrinogen receptor resulting in Glanzmann's thrombasthenia. *Blood 75:881.*
36. Maniatis, T. 1991. Mechanism of alternative pre-mRNA splicing. *Science 251:33.*
37. Higuchi, K., A. V. Hospattankar, S. W. Law, N. Meglin, J. Cortright, and H. B. Brewer, Jr. 1988. Human apolipoprotein B (apo B) mRNA: identification of two distinct apoB mRNAs, an mRNA with the apoB-100 sequence and an apoB

mRNA containing a premature in-frame translational stop codon, in both liver and intestine. *Proc. Natl. Acad. Sci. USA 85:1772.*

38. Wilson, D. B., H. H. Salem, J. S. Mruk, I. Maruyama, and P. W. Majerus. 1984. Biosynthesis of coagulation factor V by a human hepatocellular carcinoma cell line. *J. Clin. Invest. 73: 654.*
39. Cerveny, T. J., D. N. Fass, and K. G. Mann. 1984. Synthesis of coagulation factor V by cultured aortic endothelium. *Blood 63:1467.*
40. Rodgers, G. M. 1988. Vascular smooth muscle cells synthesize, secrete and express coagulation factor V. *Biochim. Biophys. Acta 968:17.*
41. Rothberger, H., and M. P. McGee. 1984. Generation of coagulation factor V activity by cultured rabbit alveolar macrophages. *J. Exp. Med. 160:1880.*
42. Chiu, H. C., P. K. Schick, and R. W. Colman. 1985. Biosynthesis of factor V in isolated guinea pig megakaryocytes. *J. Clin. Invest. 75:339.*
43. Gewirtz, A. M., M. Keefer, K. Doshi, A. E. Annamalai, H. C. Chiu, and R. W. Colman. 1986. Biology of human megakaryocyte factor V. *Blood 67:1639.*
44. Sakai, T., and W. Kisiel. 1990. Binding of human factors X and Xa to HepG2 and J82 human tumor cell lines. *J. Biol. Chem. 256:9105.*
45. Stubbs, J. D., C. Lekutis, K. L. Singer, A. Bui, D. Yuzuki, U. Srinivasan, and G. Parry. 1990. cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences. *Proc. Natl. Acad. Sci. USA 87:8417.*
46. Garcia-Sanz, J. A., H. R. MacDonald, D. E. Jenne, J. Tschopp, and M. Nabholz. 1990. Cell specificity of granzyme gene expression. *J. Immunol. 145:3111.*
47. Jenne, D. E., and T. Tschopp. 1988. Granzyme: a family of serine proteases in granules of cytolytic T lymphocytes. *Curr. Top. Microbiol. Immunol. 140:33.*
48. Kramer, M. D., and M. M. Simon. 1987. Are proteinases functional molecules of T lymphocytes? *Immunol. Today 8:140.*
49. Auberger, P., S. Sonthonnax, J.-F. Peyron, B. Mari, and M. Fehlmann. 1990. A chymotryptic-type serine protease is required for IL-2 production by Jurkat T cells. *Immunology 70:547.*
50. Gershenfeld, H. K., and I. L. Weissman. 1986. Cloning of a cDNA for a T cell-specific serine protease from a cytotoxic T lymphocyte. *Science 232:854.*

Evidence Appendix H

Short reports

Association of idiopathic venous thromboembolism with single point-mutation at Arg^{506} of factor V

Jan Voorberg, Jeanet Roelse, Rianne Koopman, Harry Büller, Fenny Berends, Jan W ten Cate, Koen Mertens, Jan A van Mourik

Abnormal coagulation factor V may underlie the thrombotic events associated with resistance to activated protein C (APC). We analysed 27 consecutive patients with documented idiopathic (recurrent) thromboembolism for the occurrence of point mutations within the APC sensitive regions of blood coagulation factor V. In 10 patients we observed a single basepair mutation resulting in a substitution of Arg^{506} to Gln. This mutation was significantly linked to in-vitro resistance to APC in these subjects. This mutation at Arg^{506} of factor V may form the molecular basis for the thrombotic events associated with APC resistance.

Lancet 1994; **343:** 1535–36

See Commentary page 1515

Venous thromboembolism has been associated with molecular defects in several haemostatic components: antithrombin III, protein C, protein S, plasminogen, and fibrinogen.[1] However, in over 90% of patients the cause remains obscure.[2] A poor anticoagulant response to activated protein C (APC) has been observed in about 20–30% of patients with an idiopathic predisposition to thromboembolic disease.[3-6] This abnormal response has been linked to a plasma factor which appeared to be identical to coagulation factor V.[7] These observations suggest that a molecular abnormality in factor V underlies the thrombotic events that are associated with a defective anticoagulant response to APC in vitro. We report linkage between resistance to APC and a single point-mutation at a putative APC cleavage site at Arg^{506} of factor V.

We investigated 27 consecutive patients (13 men; mean age 53, range 23–79) with (recurrent) idiopathic episodes of thromboembolism confirmed by contrast venography, pulmonary angiography, or both. None of the patients received oral anticoagulants at the time of the study. Patients with cancer or lupus anticoagulants were excluded. No patient had an acquired or inherited deficiency of antithrombin III, protein C, protein S, or plasminogen. Routine screening of blood coagulation and fibrinolysis revealed no abnormality. Resistance to APC was assessed by the APC-dependent prolongation of the activated partial thromboplastin time (Coatest APC Resistance, Chromogenix, Sweden).[3,5] An APC sensitivity ratio $\leq 2{\cdot}0$ was considered to represent a defect in the anticoagulant response to APC.

Patients were analysed for the presence of mutations at Arg^{506} in factor V with the following oligonucleotide primers: 5'CATCACGTTTCACCTCATCAGG3' (primer 506–2, nucleotides 1708–1730 of human factor V) and 5'ATCAGAGCAGTTCAACCAGGG3' (506–5, nucleotides 1414–1435). RNA was isolated from peripheral blood lymphocytes by the RNAzol B method (WAK Chemie, Bad Homburg, Germany) and cDNA was prepared.[8] Amplification by polymerase chain reaction with primers 506–2 and 506–5 yields a fragment of 316 basepairs, which encodes the part of factor V that contains

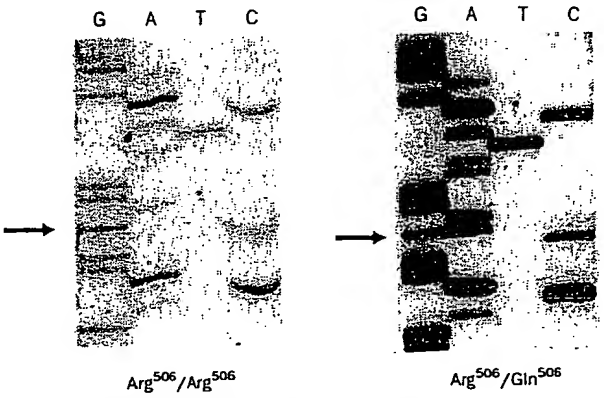

Figure 1: **Sequence analysis of factor V cDNA**

Factor V cDNA derived from patient heterozygous for Arg^{506} to Gln mutation is shown in right panel. Heterozygosity is scored by occurrence of both a "G" and an "A" at second basepair of codon Arg^{506} (CGA/CAA) of factor V (arrow). In left panel, sequence analysis of patient who does not carry the mutation is displayed. Arrow = single "G" observed at second basepair of codon Arg^{506} (CGA/CGA).

the APC cleavage site at Arg^{506}. The occurrence of mutations at Arg^{506} was monitored by direct sequencing of the amplified fragment.

Previous studies with bovine factor V have shown that APC partly inactivates factor V by cleavage at the peptide-bond Arg^{505}-Gly^{506}.[9] Direct sequencing of the corresponding part of factor V cDNA derived from our patients revealed a single G to A transition, which results in substitution of Arg^{506} for Gln (figure 1, right panel). 10 of the 27 patients were heterozygous for the Arg^{506} to Gln^{506} mutation. 8 of the 27 (30%) had an abnormal APC sensitivity ratio ($\leq 2{\cdot}0$), in agreement with the frequency in similar cohorts.[3,5] The abnormal APC sensitivity ratio was significantly linked to the Arg^{506} to Gln mutation (figure 2; *U* test, $p < 0{\cdot}0001$). In 3 patients who were heterozygous for the Arg^{506} Gln mutation, APC ratio was just above 2·0, and in only 1 patient did an abnormal APC sensitivity ratio (1·9) coincide with the normal Arg^{506}/Arg^{506} genotype.

Our results indicate that APC resistance in patients with idiopathic thromboembolism was linked to a single mutation at the putative APC cleavage site at Arg^{506} in factor V. The data suggest that APC resistance is not due to a deficiency of a cofactor of APC as has been proposed,[3,6,7,10]

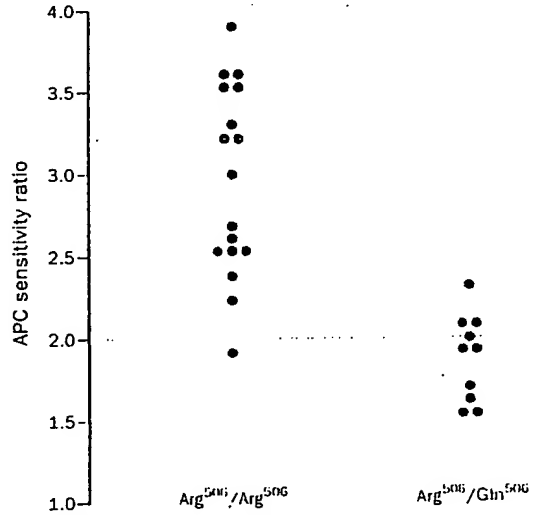

Figure 2: **Anticoagulant response to APC**

Right = 10 patients heterozygous for Arg^{506}→Gln mutation, and left = 17 patients who do not carry the mutation.

but merely reflects the inability of APC to inactivate the pro-coagulant factor V. The high frequency of APC resistance in patients with idiopathic venous thromboembolism suggests that the mutation Arg^{506} to Gln may be a major cause of inherited thrombophilia.

We thank Dr W Schaasberg and Prof W G van Aken, Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, for doing statistical analysis and for support throughout the study, respectively.

References

1 Miletich P, Prescott SM, White R, Majerus P, Bovill EG. Inherited predisposition to thrombosis. *Cell* 1993; 72: 477–80.

2 Heyboer H, Brandjes DPM, Büller HR, Sturk A, ten Cate JW. Deficiencies of coagulation-inhibiting and fibrinolytic proteins in outpatients with deep vein thrombosis. *N Engl J Med* 1990; 323: 1512–16.

3 Dahlbäck B, Carlsson M, Svensson PJ. Familial thrombophilia due to a previously unrecognized mechanism characterized by poor anticoagulant response to activated protein C: prediction of a cofactor to activated protein C. *Proc Natl Acad Sci USA* 1993; 90: 1004–08.

4 Griffin JH, Evatt B, Wideman C, Fernandez JA. Anticoagulant protein C pathway defective in majority of thrombophilic patients. *Blood* 1993; 82: 1989–93.

5 Koster T, Rosendaal FR, de Ronde F, Briët E, Vandenbroucke JP, Bertina RM. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. *Lancet* 1993; 342: 1503–06.

6 Svensson P, Dahlbäck B. Resistance to activated protein C as a basis for venous thrombosis. *N Engl J Med* 1994; 330: 517–21.

7 Dahlbäck B, Hildebrand B. Inherited resistance to activated protein C is corrected by anticoagulant cofactor activity found to be a property of factor V. *Proc Natl Acad Sci USA* 1994; 81: 1396–400.

8 Cuypers HTM, Bresters D, Winkel IN, et al. Storage conditions of blood samples and primer selection affect the yield of cDNA-polymerase chain reaction products of hepatitis C virus. *J Clin Microbiol* 1992; 30: 3220–24.

9 Kalafatis G, Mann KG. Role of the membrane in inactivation of factor Va by activated protein C. *J Biol Chem* 1993; 268: 27246–57.

10 Bauer KA. Hypercoagulability—a new cofactor in the protein C anticoagulant pathway. *N Engl J Med* 1994; 130: 566–67.

Department of Blood Coagulation, Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, Plesmanlaan 125, 1066 CX Amsterdam, Netherlands (J Voorberg PhD, J Roelse BSc, K Mertens PhD, J A van Mourik PhD); **and Centre of Haemostasis, Thrombosis, Atherosclerosis and Inflammation Research, Academic Medical Centre, Amsterdam** (R Koopman MD, H Büller MD, F Berends MD, Prof J W ten Cate MD)

Correspondence to: Dr Jan Voorberg

Linkage between inherited resistance to activated protein C and factor V gene mutation in venous thrombosis

Bengt Zöller, Björn Dahlbäck

Resistance to activated protein C (APC) is a major cause of familial thrombophilia, and can be corrected by an anticoagulant activity expressed by purified factor V. We investigated linkage between APC resistance and the factor V gene in a large kindred with familial thrombophilia. Restriction fragment length polymorphisms in exon 13 of the factor V gene were informative in 14 family members. The 100% linkage between factor V gene polymorphism and APC resistance strongly suggested a factor V gene mutation as a cause of APC resistance. A point mutation changing Arg^{506} in the APC cleavage site to a Gln was found in APC resistant individuals. These results suggest factor V gene mutation to be the most common genetic cause of thrombophilia.

Lancet 1994; **343:** 1536–38

See Commentary page 1515

Heterozygous protein C or protein S deficiency is associated with familial thrombosis,[1] and inherited resistance to activated protein C (APC) as a possible cause of thrombophilia was discovered in a family with thrombosis.[2] It is now well established that APC resistance is a major cause of venous thrombosis,[3–6] and APC resistance in different families appears to be characterised by a molecular similarity.[3] In a cohort of thrombosis patients, APC resistance was at least ten times more frequent than any of the other anticoagulant protein deficiencies (40% *vs* ≤4%) and in familial thrombophilia, it accounted for more than 50% of cases.[3] APC resistance in the general population is around 5%.[3,5]

We have isolated and characterised the protein that corrects APC resistance, and found it to be identical to factor V.[7] Factor V is pro-coagulant after activation by thrombin, whereas the novel anticoagulant cofactor activity, which we have also found in purified systems (Dr L Shen and BD, Department of Clinical Chemistry, Malmo General Hospital, Sweden), appears to be a property of unactivated factor V. Because APC-resistant plasma contains normal levels of factor V pro-coagulant activity, APC resistance may be caused by mutations in the factor V gene resulting in selective loss of the anticoagulant activity of factor V or in increased resistance to APC of mutant factor Va itself. We have investigated whether APC resistance is due to mutation in the factor V gene in a study of linkage in a large kindred with familial thrombophilia.

The APC resistance test, a modified activated partial thromboplastin time in which the anticoagulant response to standardised addition of APC is measured, was done as described.[2,3] The results were expressed as the APC ratio (clotting time with the APC/$CaCl_2$ solution divided by clotting time with $CaCl_2$). Family members with confirmed APC ratios under 2·0 were considered to be APC resistant.[3] Free and total protein S were measured with a radioimmunoassay.[3] Family members with a concentration of free protein S below the normal limit were considered to be protein S deficient. Their total protein S levels were slightly low or in the low normal range.

Genomic DNA was prepared from EDTA-blood by standard procedures. A sequence of 1188 basepairs (bp) of the factor V gene (nucleotides 2066 to 3254 of the factor V cDNA; sequence from Genebank) was amplified from genomic DNA with two primers 5'GAACTTGGATGTTAACTTCC3' and 5'GGCTTCACT-TCTTAGAGGGTG3' (figure). The conditions for polymerase chain reaction (PCR) for 40 cycles of amplification were: 60 s denaturation at 93°C, 30 s annealing at 61°C, and 180 s extension at 72°C. After amplification, the DNA was cleaved with *Taq*1 and with *Eco*R1, and subjected to agarose-gel electrophoresis. The region in exon 10 that encodes one of the APC cleavage sites in factor V was PCR amplified from genomic DNA with 5'GGG-CTAATAGGACTACTTCTAATC3' (corresponding to Gly^{490}-Ile^{497}) and 5'TCTCTTGAAGGAAATGCCCCATTA3' (derived from intron sequence provided by Dr W Kane). The PCR conditions were 5 min initial denaturation at 94°C followed by 30 cycles of 60 s denaturation at 93°C, 30 s annealing at 61°C, and 90 s

RELATED PROCEEDINGS APPENDIX

None